

Efficacy of noisy dynamical decoupling

Jiaan Qi,¹ Xiansong Xu,² Dario Poletti,² and Hui Khoon Ng^{1,3,4}

¹*Yale-NUS College, Singapore*

²*Singapore University of Technology and Design*

³*Centre for Quantum Technologies, National University of Singapore*

⁴*MajuLab, International Joint Research Unit UMI 3654, CNRS-UCA-SU-NUS-NTU, Singapore*

(To be edited.) This paper study dynamical decoupling (DD), a family of pulse schemes aimed at removing noise in quantum systems. We first review the formalism of DD with an emphasis on the error phase, a figure-of-merit characterizing the performance of DD. We then develop noise suppression criteria from the error phase perspective for the periodic and concatenated DD schemes. A rigorous proof for the noise decoupling order of concatenated DD is also presented.

I. INTRODUCTION

The adverse effects of noise remains the single biggest obstacle to the realisation of large-scale quantum technologies. The very quantum effects that give quantum technologies their edge over classical devices are extremely fragile and easily destroyed by the presence of unwanted interactions with the environment—noise. Much of the current research and technological push in the community today are centered around implementing methods to remove the effects of noise in quantum devices.

A key noise-removal technique is dynamical decoupling (DD) [1–11], a low-resource-cost approach that only requires application of fast pulse sequences on individual qubits in the quantum device to average away the effects of noise processes slow compared to the pulse time. With its roots in spin-echo techniques in NMR systems, DD has been used in many different types of experiments as a simple way to reduce noise in quantum information processing systems. Such active noise removal scheme is in direct contrast with the widely studied approach of quantum error correction (QEC) [12]. Compared with QEC, dynamical decoupling is typically more economical as it require no encoding of logical qubits using multiple physical qubits, nor real-time close-loop control through periodic syndrome measurement. All that is needed are regular single-qubit fast pulses that are usually easy to implement. DD be used by itself, or as the first layer of defense against noise In practice, one can choose to only implement DD or to incorporate DD within a standard QEC scheme as the low-level noise-reduction approach.

The use of DD does not, however, come at no cost. To implement it, one has to apply multiple pulses, which can be imperfect. On the one hand, if the DD pulses are perfect, the slow noise will be averaged away, leaving weaker residual noise on the system; on the other hand, if the DD pulses are imperfect, those imperfections can add errors to the system, and if those errors happen often enough, they can eliminate the benefit of having DD in the first place. Like the cost-benefit analysis familiar to fault-tolerant quantum computing, one has to ask whether there is an error threshold for DD, limiting the level of imperfections allowed in the DD pulses above

which DD simply cannot offer any benefit.

More to come.

II. PRELIMINARIES

A. DD basics

Dynamical decoupling (DD) involves the repeated application of a fixed sequence of short pulses (or fast gates) to individual quantum registers—usually qubits—that averages away the effect of any noise with a time scale slow compared to the sequence time.

A DD scheme is essentially a recipe specifying how a particular sequence of pulses should be applied. This consists of an ordered list of gates (P_1, P_2, \dots, P_L), together with their application time (t_1, t_2, \dots, t_L), which is assumed to lie within the time interval $[t_0, t_L]$. All DD schemes share some fundamental similarities. First, the gates are chosen from a specified group of transformations $P_i \in \mathcal{G}$, $\forall i$. Second, any DD sequence satisfies the constraint that the gates product to identity:

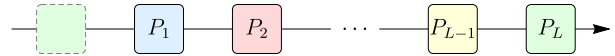
$$P_L P_{L-1} \dots P_1 \doteq I, \quad (1)$$

where the operators proceeds from right to left in time; the dot above the equal sign means “up to an overall phase factor”. This constraint ensures that there is no net transformation on the quantum register at the completion of the sequence. Different strategies could differ in the following aspects: (i) the transformation group \mathcal{G} .—For a qubit register, an obvious choice is perhaps the Pauli group generated by the Pauli operators X, Y , and Z . Simpler schemes such as spin echo [13] and the CPMG sequence [14, 15] use only the subgroup $\{I, Z\}$. (ii) the sequence length L and the specific sequence of pulses. (iii) the pulse times.—One can have regular-interval pulses, with $t_{i+1} - t_i$ remaining constant for all i , as is the case in the periodic DD (PDD) [1] and concatenated DD (CDD) [4] schemes. One could, however, have variable-interval schemes, such as the Uhrig DD (UDD) sequence [7], nested UDD [9], and quadratic DD [11].

In this work, we focus on schemes of PDD and CDD, two closed related schemes that both employ gates drawn from the Pauli group and have regular gate intervals. The restriction to regular-interval schemes is of practical relevance. Typically there is no way to place two pulses arbitrary close to each other given technological constraints. Also in many cases quantum gates are encapsulated as black boxes whose implementation cannot be altered. Hence the overall physical picture here is a quantum circuit working under fixed frequency, with time τ in between consecutive gates. A nonzero τ can be simply the finite switch time between gates, or there may be other practical reasons for a synchronized clock cycle time. The control gates within a DD sequence are modeled as sharp pulses. We illustrate a typical sequence of DD control gates in Figure 1.

There are two sources of errors: (i) the system—the quantum register—interacts with a bath, and that interaction is the source of noise on the system, present even in the absence of the DD pulses; (ii) the pulse imper-

circuit:



model:

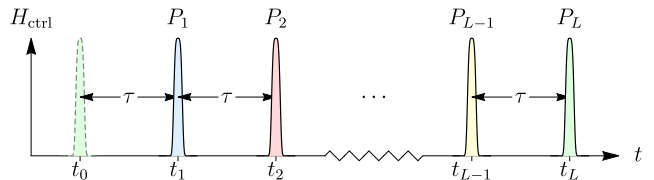


FIG. 1: Illustration of a typical DD sequence. In the diagram, different control gates/pulses are filled with different colors. The imaginary gate (dashed) at t_0 corresponds to the last gate from the previous cycle.

fections. In past work (see, for example, Ref. [4]), imperfect DD pulses were modeled as finite-duration pulses during which the always-on system-bath interaction acts and leads to errors in the pulses. Here, we allow for the more general, and practically relevant, situation of additional control errors in the pulses. We model noisy finite-duration pulses as instantaneous ideal pulses followed by a noise map that captures the background noise as well as any additional control errors. DD sequences are ideally designed to average away the background noise arising from the system-bath interaction. We refer to the original DD sequences with perfect instantaneous pulses as ideal DD, and to the case with imperfect pulses as noisy DD.

In the absence of any DD pulses, the system and bath evolve jointly according to the Hamilton operator H , assumed to be time-independent as is appropriate for standard DD analysis (put differently, any parametric time dependence in the joint dynamics occurs because of degrees of freedom excluded from the bath; here we think of all such degrees of freedom as part of the bath). H here can be written as $H = H_B + H_{SB}$, with H_B as the bath-only Hamilton operator, and H_{SB} the interaction. No system-only term appears in H as we assume no nontrivial dynamics (other than that arising from H_{SB}) occur in the system during the DD period. With this, we can write the evolution operator for ideal DD as

$$U_{DD} \equiv P_L e^{-i\tau H} P_{L-1} \dots P_2 e^{-i\tau H} P_1 e^{-i\tau H} \equiv e^{-i\Omega_{DD}}. \quad (2)$$

Here, we have defined $\Omega_{DD} \equiv T H_{\text{eff}}$ as the dimensionless (with the T factor) effective Hamilton operator appropriate for describing evolution for the time of the DD sequence. Ω_{DD} can be written down formally using the Magnus expansion,

$$\Omega_{DD} = \sum_{m=1}^{\infty} \Omega_{DD}^{(m)}, \quad (3)$$

where the m th term consists of products of m copies of τH , and hence of order $\|\tau H\|^m$. We refer the reader

to past analyses of DD (cite) for a detailed derivation; here we provide only the basic expressions needed for our discussion below. (Add back a bit of the derivation of the $\Omega_{\text{DD}}^{(1)}$ and $\Omega_{\text{DD}}^{(2)}$ terms.) Within the radius of convergence of the Magnus series, $\Omega_{\text{DD}}^{(m)}$ decreases in importance as m increases. A DD scheme such that $\Omega_{\text{DD}}^{(m)}$ acts trivially on the system (i.e., acts as I on system) for all $m \leq n$ is said to achieve n th-order decoupling—the system sees weakened noise, of strength $\|\Omega_{\text{DD}}^{(n)}\| \sim \|\tau H\|^n$, compared to $\|\Omega_{\text{DD}}^{(1)}\| \sim \|\tau H\|$ without DD.

B. Quantifying the efficacy of DD

1. Decoupling Order

According to Eq. (??), the DD protocol can be regarded as a transformation $\Omega \rightarrow \Omega_{\text{DD}}$. Since the noise map \mathcal{E} is a function of Ω and ρ_{B} , the transformation on Ω takes the original quantum channel into a new one $\mathcal{E} \rightarrow \mathcal{E}_{\text{DD}}$. As is argued in Sec. ??, error phase is a convenient characterization of noise strength, we apply the two-component norm to the dimensionless Hamiltonian before and after DD, leading to

$$\|\Omega\| \equiv (\alpha, \beta) \quad \text{and} \quad \|\Omega_{\text{DD}}\| \equiv (\alpha_{\text{DD}}, \beta_{\text{DD}}). \quad (4)$$

The “decoupling” reference in the name of DD is understood in terms of the coupling part: i.e., the error phase of the new channel gets reduced through this transformation. We say noise is suppressed if the ratio $\beta_{\text{DD}}/\beta < 1$, and a lower ratio would indicate better noise suppression.

A widely used indicator to characterize the performance of DD schemes is the *decoupling order*. A DD scheme is said to achieve n th-order decoupling if the first n terms in the series (3) have vanishing coupling part:

$$\mathcal{P}_{\text{c}}(\Omega_{\text{DD}}^{(k)}) = 0, \quad \text{for } 1 \leq k \leq n. \quad (5)$$

The decoupling order is thus a single positive integer to grade different DD schemes. Greater decoupling order is usually linked with better noise suppression rate. This can be seen by evaluating the norm of the Magnus series projected on to the coupling part. For an n th-order decoupling sequence, we have

$$\beta_{\text{DD}} = \left\| \sum_{k=n+1}^{\infty} \mathcal{P}_{\text{c}}(\Omega_{\text{DD}}^{(k)}) \right\| \simeq (\alpha, \beta)^{n+1}, \quad (6)$$

where we use the notation $(\alpha, \beta)^{n+1}$ to represent a polynomial that is $(n+1)$ th order in α and β . Provided that the Magnus series converges absolutely, β_{DD} is bounded by a power series in $\|\Omega\|$ whose leading term is on the order of $n+1$. Therefore Eq. (6) provides an alternative characterization for the n th-order decoupling condition, which can be useful when directly calculating the first n Magnus terms is unfeasible.

2. Error Phase

To gauge the efficacy of DD, we need to quantify the deviation of the actual state of the quantum system, with and without DD, from the ideal, no-noise state. Following Ref. [5], we make use of the *error phase*, which measures the strength of the system-bath interaction, the source of the noise on the system. The system and bath evolve jointly for some specified time T according to the evolution operator $U(0, T)$. The underlying joint Hamilton operator generating the dynamics can be time-dependent, and can include—or not—the DD pulses on the system. We write $U(0, T) \equiv e^{-iTH_{\text{eff}}}$, for some effective time-independent system-bath Hamilton operator H_{eff} . H_{eff} can be split into two pieces: $H_{\text{eff}} \equiv H_{\text{eff,B}} + H_{\text{eff,SB}}$, where $H_{\text{eff,B}} \equiv \frac{I_{\text{S}}}{d_{\text{S}}} \otimes \text{tr}_{\text{S}}(H_{\text{eff}})$ acts on the bath alone, while $H_{\text{eff,SB}} \equiv H_{\text{eff}} - H_{\text{eff,B}}$ contains all the pieces that act nontrivially on the system. $H_{\text{eff,SB}}$ can be thought of as the effective system-bath interaction over this time T . We define the error phase as the norm of that interaction, multiplied by the time interval,

$$\Phi_{\text{SB}} \equiv \|TH_{\text{eff,SB}}\|. \quad (7)$$

We use a norm such that $\|S \otimes B\| = \|S\| \|B\|$, and $\|I\| = 1 = \|X\| = \|Y\| = \|Z\|$ for single-qubit I, X, Y , and Z .

As we will see, the norm of $H_{\text{eff,B}}$ will also enter our analysis. We define the corresponding error phase for the bath-only term as $\Phi_{\text{B}} \equiv \|TH_{\text{eff,B}}\|$, and also “two-component” error phase as

$$\Phi \equiv \|\|TH_{\text{eff}}\|\| \equiv (\Phi_{\text{B}}, \Phi_{\text{SB}}) \quad (8)$$

(To merge with new text.) Following the properties of operator norm, thus-defined map is absolutely-scalable $\|\|a\Omega\|\| = |a| \cdot \|\|\Omega\|\|, \forall a \in \mathbb{C}$, and satisfies the inequality, $\|\|A+B\|\| \leq \|\|A\|\| + \|\|B\|\|$, where the “ \leq ” sign between number pairs is defined by each of the components: $(a_1, b_1) \leq (a_2, b_2)$ if both $a_1 \leq a_2$ and $b_1 \leq b_2$; vice versa for the “ \geq ” sign. It can also be easily verified that the two-component norm is invariant under arbitrary disjoint unitary transformation on $\mathcal{H}_{\text{S}} \otimes \mathcal{H}_{\text{B}}$:

$$\|\|(U_{\text{S}} \otimes U_{\text{B}})\Omega(U_{\text{S}}^{\dagger} \otimes U_{\text{B}})\|\| = \|\|\Omega\|\|, \quad \forall U_{\text{S}}(U_{\text{B}}) \in \mathcal{U}(d_{\text{S(B)}}), \quad (9)$$

where $\mathcal{U}(d_{\text{S(B)}})$ is set of unitary operators on \mathcal{H}_{S} or \mathcal{H}_{B} . This invariance suggests that the error phase is invariant under arbitrary interaction picture defined by a system or bath evolution.

C. PDD

We specialize here to the case of interest, that of the single-qubit PDD scheme. The single qubit interacts with a bath, with a joint Hamilton operator (in the absence of DD) that can be written, without loss of generality, as

$$H \equiv I \otimes B_I + X \otimes B_X + Y \otimes B_Y + Z \otimes B_Z, \quad (10)$$

where I is the identity operator on the qubit, X, Y , and Z are the Pauli operators on the qubit, and B_i , for $i = I, X, Y, Z$, are operators on the bath. We identify $I \otimes B_I$ as the bath-only H_B and $X \otimes B_X + Y \otimes B_Y + Z \otimes B_Z$ as the interaction Hamilton operator H_{SB} , and define

$$\|\tau H\| = (\|\tau H_B\|, \|\tau H_{SB}\|) \equiv (\phi_B, \phi_{SB}). \quad (11)$$

(I moved τ inside the norm operator to conform the convention that norm should be taken on a dimensionless operator, not on Hamiltonian alone. But in the coming texts there are many violations.)

The PDD scheme uses a simple 4-pulse sequence $ZXZX$. The corresponding Ω_{PDD} of Eqs. (2) and (3), now re-labeled as Ω_{PDD} , can be worked out to be

$$\Omega_{\text{PDD}} = \Omega_{\text{PDD}}^{(1)} + \Omega_{\text{PDD}}^{(2)} + \dots \quad (12)$$

$$\text{with } \Omega_{\text{PDD}}^{(1)} = (4\tau)I \otimes B_I \quad (13)$$

$$\Omega_{\text{PDD}}^{(2)} = -(2\tau^2)\{X \otimes 2i[B_I, B_X] + Y \otimes (i[B_I, B_Y] + \{B_X, B_Z\})\}, \quad (14)$$

where $[\cdot, \cdot]$ and $\{\cdot, \cdot\}$ are the commutator and anti-commutator, respectively. Ω_{PDD} should be compared against the evolution operator (over the same time period of 4τ) in the absence of DD: $U = e^{-i4\tau H} \equiv e^{-i\Omega}$, with

$$\Omega = (4\tau)H = (4\tau)(I \otimes B_I + H_{SB}) \quad (15)$$

Comparing $\Omega_{\text{PDD}}^{(1)}$ —usually the dominant term—with Ω , we see that the H_{SB} in Ω no longer appears in $\Omega_{\text{PDD}}^{(1)}$ and $\Omega_{\text{PDD}}^{(1)}$ is trivial on the system. This corresponds to the fact that PDD is able to remove the lowest-order noise and that it achieves first-order decoupling. (To revisit to see if a re-phrasing is needed. I'm worried about the $[B_I, B_X]$ term in $\Omega_{\text{PDD}}^{(2)}$, which is of order $\|H_B H_{SB}\|$, not $\|H_{SB}\|^2$. Also the last sentence of the paragraph following Eq. (3).)

[I am changing most of the Ω -like quantities into τH s (except when we talk about the Magnus series) – it is important to keep the time parameter visible. Sometimes, τ is the relevant time—for the no DD case—sometimes $K\tau$ is the relevant one—for the DD case. These should just be visible in the analysis, not kept hidden within Ω . This becomes especially important when considering CDD.]

D. Scaling up protection

A given DD sequence yields a given decoupling order, setting a limit on the scheme's ability to reduce noise in the system. To increase the power of the DD scheme, one can employ the method of concatenation introduced in Ref. [4]. In that work, concatenated DD (CDD) was built upon the basic PDD scheme; the same procedure of concatenation, however, can be applied to other basic

DD sequences. The idea is to make use of concatenation to increase the decoupling order of the resulting DD sequence, hence scaling up the protection of the system against noise.

CDD can be described in a recursive manner. We begin with the bare evolution, without any DD sequence, writing the evolution operator over some time interval t as

$$U_0(t) \equiv e^{-itH_0}, \quad (16)$$

where $H_0 \equiv H$, with the subscript 0 added here in preparation for concatenation to higher levels. With a DD scheme of L pulses, applied at time intervals (assuming regular-interval DD) τ_0 , the evolution operator is

$$U_1(\tau_1) \equiv P_L U_0(\tau_0) \dots P_2 U_0(\tau_0) P_1 U_0(\tau_0), \quad (17)$$

where $\tau_1 \equiv L\tau_0$. The subscript 1 is to be understood as indicating that this is concatenation level 1. To concatenate further, $U_k(\tau_k)$ is defined recursively,

$$U_{k+1}(\tau_{k+1}) \equiv P_L U_k(\tau_k) \dots P_2 U_k(\tau_k) P_1 U_k(\tau_k), \quad (18)$$

increasing the concatenation level by 1 each time, and with $\tau_{k+1} \equiv L\tau_k$. With each $U_k(\tau_k)$, we associate an effective Hamilton operator H_k , and a dimensionless Hamilton operator $\Omega_k \equiv \tau_k H_k$, such that

$$U_k(\tau_k) \equiv e^{-i\tau_k H_k} = e^{-i\Omega_k}. \quad (19)$$

Each CDD scheme is determined by specifying the maximal concatenation level n , and either the value of τ_0 or τ_n . CDD at level n , denoted as CDD_n , is then a sequence of $L_n \equiv L^n$ pulses separated by time interval τ_0 and taking total time $\tau_n = L^n \tau_0$ to complete.

In the remainder of the paper, we will restrict our discussion to CDD built upon the basic PDD scheme. The authors of Ref. [4] showed that CDD_n achieves n th order decoupling. This quantifies the benefit of scaling up the noise protection by concatenation. Appendix B re-derives this conclusion with a different and what we believe to be clearer argument than that in the original reference.

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

$$\text{Memory setting: } \epsilon_{\text{DD}} < \epsilon(L), \quad (20)$$

where ϵ_{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

$$\text{Computational setting: } \epsilon_{\text{DD}} < \epsilon(1), \quad (21)$$

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

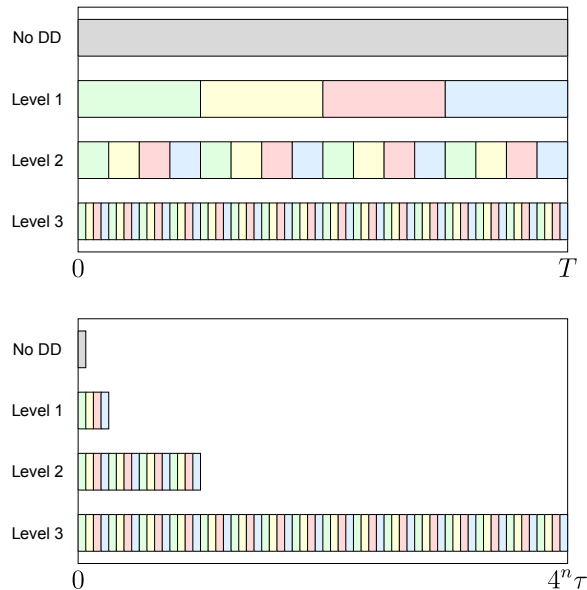


FIG. 2: An illustration of the CDD time evolution maps under increasing concatenation levels. We compare the quantum memory setting (above) to the quantum computation setting (under).

These two perspectives result in very different overall decoupling strategies. Especially if one is using CDD to “scale up protection”. As illustrated in Fig. 2, 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 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 (21) 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 Φ_{SB} , not our two-parameter Φ_{SB} and η situation. In fact, in current experiments η 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 2). 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 (α, β) values would lead 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.

III. LIMITS OF DECOUPLING – THE BREAK-EVEN CONDITIONS

Now we have introduced the basic principles of DD, a natural question will be how

Any noise mitigation strategy is effective only if its costs—the increased complexity of carrying out the quantum computational task—are lower than its benefits—the increased ability to remove adverse effects of the noise. The costs enter not just because any noise mitigation approach requires the use of more resources, e.g., more gates, more qubits, etc., but more that those added resources are themselves imperfect in practice, such that the unmitigated noise of the resulting larger system is unavoidably larger than if no noise mitigation strategy was adopted. There is then a net benefit only if the added noise is low enough to not overwhelm the added noise-removal capabilities. Such is the content of any fault-tolerant analysis, usually applied to quantum computing tasks protected by quantum error correction. Here, we apply the same logic to DD, and ask when the added benefit of averaging away part of the noise outweighs the added cost of having to do additional DD pulses.

A. Ideal case

Let us first discuss when ideal PDD is useful—we expect to recover the usual statement of when the idea of PDD works at all. We choose the error phase as our figure of merit: $\epsilon \equiv \Phi$. For PDD to be useful, we require, under the computational setting,

$$\Phi_{\text{SB}} \leq \phi_{\text{SB}}. \quad (22)$$

From our earlier discussion of PDD, it is clear that $\Phi_{\text{SB}} = \|\Omega_{\text{PDD}}^{(2)} + \Omega_{\text{PDD}}^{(3)} + \dots\|$. The full Magnus series is difficult to write down, but we can employ the dominant term, $\Omega_{\text{PDD}}^{(2)}$ which acts nontrivially on the system, for the approximate condition:

$$\Phi_{\text{SB}} \simeq \|\Omega_{\text{PDD}}^{(2)}\| \leq \phi_{\text{SB}}. \quad (23)$$

Using the expression for $\Omega_{\text{PDD}}^{(2)}$ from Eq. (??), we have

$$\begin{aligned} \|\Omega_{\text{PDD}}^{(2)}\| &\leq 2\tau^2(2\|[B_I, B_x]\| + \|[B_I, B_Y]\| + \|\{B_X, B_Z\}\|) \\ &\leq 4\phi_B\tau(2\|B_X\| + \|B_Y\|) + 4\tau^2\|B_X\|\|B_Z\| \\ &\leq 12\phi_B\phi_{\text{SB}} + 4\phi_{\text{SB}}^2 \end{aligned} \quad (24)$$

noting that $\phi_B \equiv \tau\|H_B\| = \tau\|B_I\|$, and that $\phi_{\text{SB}} \equiv \tau\|H_{\text{SB}}\| = \tau\left[\sum_{i=X,Y,Z}\|B_i\|^2\right]^{1/2} \geq \tau\|B_i\|$ for any $i = X, Y, Z$ (The first equality is true under the H-S induced norm, but missing a factor of 2, as written; can be remedied by using normalized Pauli operators. Need to fix this throughout the paper). To get condition (23), it then suffices to require

$$12\phi_B + 4\phi_{\text{SB}} \leq 1. \quad (25)$$

This amounts to a requirement that both ϕ_{SB} and ϕ_B be small for PDD to work well. That ϕ_{SB} has to be small comes as no surprise— ϕ_{SB} quantifies the strength of the noise on the system, and DD is expected to work well as long as the noise is weak enough so that the remnant noise is weak. The requirement that ϕ_B be small is perhaps a bit more surprising – after all, it bounds the bath-only term which does not directly lead to noise on the system. Nevertheless, H_B determines the evolution rate of the bath. The requirement that ϕ_B also has to be small should be understood as the requirement that any DD scheme requires the noise to remain more or less constant throughout the sequence, for good averaging by the pulses. ϕ_B quantifies how rapidly the bath evolves, and hence how quickly the noise seen by the system changes (through the evolution of the bath operators B_i in the Heisenberg picture). A slowly evolving noise requires then a small ϕ_B . Tracing back to the original calculation, we see that part of the pre-factor of 12 of ϕ_B is the “4” for the length of the PDD sequence; a longer sequence will have a larger pre-factor and hence a more stringent requirement that ϕ_B is small, consistent with our interpretation here and our standard understanding of DD.

(To be edited and merged with the new text.) On the other hand, the inequalities (??) inspires use to use the following approximation for the error phase: (The following is exactly true, not just approximately so, up to these factors of 2 that can be fixed – but don't matter for the final inequality for the threshold condition – for the Hilbert-Schmidt (H-S) induced norm. What is the norm used in producing Fig. 3? If that was the H-S norm, that would explain why this version of the inequality works well.) (That is the infinite norm, i.e., the largest eigenvalue, and consistently used across the original script. The reason was explained in the original error phase section. Basically the norm need to satisfy $\|I\| = 1$ for whatever dimensionality of the identity. This is necessary for attaching an ancilla system without changing the error phase.—Jiaan)

$$\left\|\sum_{i=1}^3 \sigma_i \otimes B_i\right\| \approx \sqrt{\sum_{i=1}^3 \|B_i\|^2}. \quad (26)$$

Such approximation is valid up to a factor $\in [1/\sqrt{3}, \sqrt{3}]$. This approximation allows us to derive an “approximate condition”. We estimate

$$\begin{aligned} \beta_{\text{PDD}}^2 &\approx \|-4i[B_0, B_1]\|^2 + \|-2i[B_0, B_2] - 2\{B_1, B_3\}\|^2 \\ &\leq (8\alpha\|B_1\|)^2 + (4\alpha\|B_2\| + 4\|B_1\|\|B_3\|)^2, \end{aligned} \quad (27)$$

Defining $x_i \equiv \|B_i\|/\beta$, then upper-bounding the r.h.s. by β^2 requires

$$\max_{x_1^2+x_2^2+x_3^2=1} (2\alpha x_1)^2 + (\alpha x_2 + \beta x_1 x_3)^2 \leq \frac{1}{16}. \quad (28)$$

Solution of this constrained optimization problem leads

to:

$$\left(\frac{\beta}{\sqrt{3}} \leq \alpha \leq \frac{1}{8}\right) \text{ or } \left(10\alpha^2 + \frac{9\alpha^4}{\beta^2} + \beta^2 \leq \frac{1}{4}\right). \quad (29)$$

This set of inequalities automatically implies (??), it gives a tighter but less strict condition due to the approximation used for the error phase.

The criteria (??) and (29) each determines a region in the (α, β) parameter space where noise suppressing is achieved using the leading order Magnus term. But in order for the calculations to make sense, we have assumed that the Magnus series converge absolutely. According to (??), the necessary condition here is $4\|\Omega\|_\infty < \xi$. Since $\|\Omega\| \leq \alpha + \beta$, we may use a more stringent convergence bound:

$$\alpha + \beta \leq \frac{1}{4}. \quad (30)$$

To examine the accuracy of the two theoretical bounds, we perform numeric tests as a verification. We first randomly generate four 2×2 Hermitian matrices for the bath operators satisfying the normalization condition in Eq. (??), then simulate the PDD protocol and use matrix logarithm to find the actual Ω_{PDD} and error phase β_{PDD} . For each fixed (α, β) point, we numerically maximize the error phase reduction ratio β_{PDD}/β over 1000 random instances. The result is presented in Figure 3, where a point in blue color represents a greater-than-one ratio and yellow color represents smaller-than-one ratio. Thus the graph reflects the true noise suppressing condition by the continuous yellow-colored region. We also plot the boundaries for the two theoretical noise suppression region from (??) and (29), together with the absolute convergence region (30) for comparison.

As long as the two-component norm of the free evolution Hamiltonian falls in the yellow region on the lower-left corner, the idealist PDD scheme is guaranteed to work, in the sense of reducing the error phase. This reflects the fundamental “small noise” assumption of DD. From the numerical results, we learn that the bound by the “strict condition” from (??) is too loose compared with the “approximate condition” (29), which works surprisingly well in reflecting the true boundary of the noise-reduction region. The faithfulness of the latter bound is particularly striking considering the points beyond the absolute-convergence boundary (above the green dotted line), where our derivation should be unreliable in principle. An explanation is that the absolute-convergence criteria (30), sufficient but not necessary, is a loose one and the leading order Magnus series works well beyond the region predicted by this inequality. This piece of evidence boosts our confidence in using the leading order Magnus term as approximation in the following parts of this work.

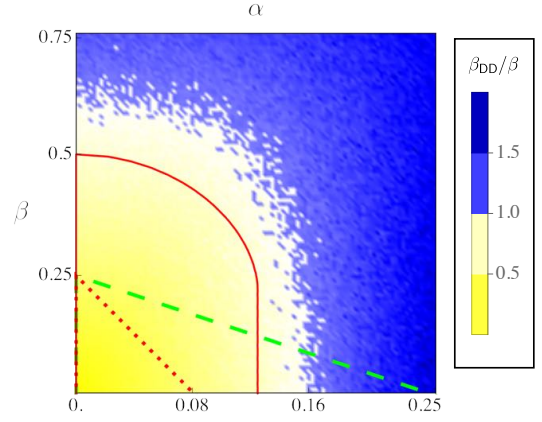


FIG. 3: The noise reduction ratio β_{DD}/β obtained through maximization over random bath operators plotted in the (α, β) parameter plane. In comparison, the theoretical boundaries for noise reduction are also plotted, with dotted red line for (??) and solid red line for (29), also with the boundary for the absolute-convergence condition (30) in dashed green line. Note the ranges for α and β are chosen differently for better visual effect.

B. Noisy gates

Previously, we treated the DD control gates in an ideal manner, which implies instantaneous and noise-free pulses. In practice, of course, the control gates are neither instantaneous nor noise-free. Hence the analysis for ideal DD cannot be directly carried out.

The effects of quantum noise on the system can be generally modelled with CPTP channels, which need not be unitary. The introduction of non-unitarity is problematic as the error phase, which is used to study the efficacy of DD, can no longer be defined. However, one may include some ancillary qubits to transform any quantum channel into a unitary evolution over the extended Hilbert space. For the analytical treatment below, we will always assume this dilation for the noisy gates. Hence, we can ascribe Hamiltonian generators for the noisy control gates. Such generator can be time-dependent and involves in the term necessary for the control in addition to various sources contributing to the noise. Focusing on the i th gate-interval, we can write the total Hamiltonian:

$$H(t) = H_e + H_c(t), \quad t_{i-1} < t \leq t_i, \quad (31)$$

where $H_e = H_B + H_{\text{SB}}$ is the error Hamiltonian between system and bath, $H_c(t)$ is the control Hamiltonian generating the gate P_i . For the ideal case considered earlier, it is assumed $H_c(t) \propto \delta(t - t_i)$; then the corresponding time evolution operator is simply $P_i e^{-i\tau H_e} = P_i e^{-i\Omega}$. For the case of noisy gates, the control Hamiltonian $H_c(t)$ is not a delta spike and will only approximately generate the gate P_i .

In general, directly solving the DD time evolution op-

erator $U(t_L, t_0)$ from $H(t)$ in Eq. (31) will not work, as the very large $H_c(t)$ integral in the matrix exponent will break the convergence condition for the Magnus series. Hence it is necessary to separate out the “pure gate” parts from the “free evolution” parts. We formally perform such separation by expressing the solution to unitary time evolution for the i th gate interval as:

$$\tilde{U}(t_i, t_{i-1}) = P_i e^{-i\tilde{O}_i}, \quad (32)$$

where the gate-dependent \tilde{O}_i replaces the free evolution generator Ω for ideal DD. We put tildes over operators to signify that the DD control gates are noisy. Similar to the ideal DD case, we can split the P_i s by defining $P_i = G_{i+1}^\dagger G_i$ and then,

$$\tilde{U}_{\text{DD}} = e^{-iG_L \tilde{\Omega}_L G_L^\dagger} \dots e^{-iG_2 \tilde{\Omega}_2 G_2^\dagger} e^{-iG_1 \tilde{\Omega}_1 G_1} \equiv e^{-i\tilde{\Omega}_{\text{DD}}}. \quad (33)$$

Provided the noise being small, the $\tilde{\Omega}_i$ s are small and we may use the Magnus formula to derive the leading order Magnus series as good approximation to $\tilde{\Omega}_{\text{DD}}$. The overall goal is to figure out the conditions such that the noise-inflicted DD error phase $\tilde{\Phi}_{\text{SB}} \equiv \|\tilde{\Omega}_{\text{DD,SB}}\|$ is still reduced compared to the bare Hamiltonian error phase $\phi_{\text{SB}} = \|\Omega_{\text{SB}}\|$. For simplicity, let us employ a single parameter η , which is defined later, to quantify the noise strength associated with the gates. The break-even condition is, in essence, to invert the inequality

$$\tilde{\Phi}_{\text{SB}}(\phi_B, \phi_{\text{SB}}, \eta) < \phi_{\text{SB}}, \quad (34)$$

and obtain an upper bound on the noise level $\eta < \eta_{\text{max}}(\phi_B, \phi_{\text{SB}})$. To achieve this, we not only need an appropriate noise strength parameter η , but also correctly estimate the DD error phase function $\tilde{\Phi}_{\text{SB}}(\phi_B, \phi_{\text{SB}}, \eta)$.

In the followings, we will mainly focus on the PDD to derive explicit break even conditions for various noise. But our analysis should be easily generalized to other DD sequences as well.

1. Noisy pulses

If the non-zero duration of $H_c(t)$ is much shorter than the gate interval, the control gates can still be viewed as instantaneous pulses. In this section we consider noisy pulses, that is $H_c(t) \sim \delta(t - t_i)$, but does not produce the ideal gate. The time evolution operator for the i th gate interval becomes:

$$\tilde{U}(t_i, t_{i-1}) = \tilde{P}_i e^{-i\Omega}, \quad (35)$$

where \tilde{P}_i is the noisy pulse version for the ideal P_i gate and $\Omega = \tau H_e$ is the free evolution generator. Let us also associate to \tilde{P}_i a Hermitian generator, which should be dominated by that of an ideal gate in addition to a small noise part:

$$\tilde{P}_i = \exp[-i(\Omega_{P_i} + \eta \Gamma_{P_i})], \quad (36)$$

where Ω_{P_i} is responsible for the ideal gate $P_i = \exp(-i\Omega_{P_i})$ and supported only on the system; and $\eta \Gamma_{P_i}$ is responsible for the noise and supported on the system-bath-ancilla composite in general. We have introduced the explicit small parameter η to facilitate order tracking and normalized Γ_{P_i} . For PDD in particular, the sequence comprises only X and Z gates. And we may write the noisy X and Z pulses as:

$$\tilde{X} = e^{-i(\frac{\pi}{2}\sigma_1 + \eta \Gamma')}, \quad \tilde{Z} = e^{-i(\frac{\pi}{2}\sigma_3 + \eta \Gamma'')}, \quad (37)$$

where Γ' and Γ'' accounts for the noise in \tilde{X} and \tilde{Z} . We separate out the system part of these operators by $\Gamma' = \sum_i \sigma_i \otimes B'_i$ and $\Gamma'' = \sum_i \sigma_i \otimes B''_i$, with normalized operators B'_i and B''_i acting on the bath-ancilla product space. The time evolution for the noisy PDD sequence then becomes

$$\begin{aligned} \tilde{U}_{\text{DD}} &= \tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \\ &= e^{-i\Gamma_3} e^{-i\Omega_3} e^{-i\Gamma_2} e^{-i\Omega_2} e^{-i\Gamma_1} e^{-i\Omega_1} e^{-i\Gamma_0} e^{-i\Omega_0}, \end{aligned} \quad (38)$$

where $\Omega_i \equiv \sigma_i \Omega \sigma_i$ as in the ideal case; additionally we introduce $e^{-i\Gamma_3} \equiv \tilde{Z} Z$, $e^{-i\Gamma_2} \equiv Z \tilde{X} Y$, $e^{-i\Gamma_1} \equiv Y \tilde{Z} X$, and $e^{-i\Gamma_0} \equiv X \tilde{X}$. By keeping the first order in η , we can derive

$$\begin{aligned} \Gamma_3 &= \eta(\sigma_0 \otimes B''_0 - \frac{2}{\pi} \sigma_1 \otimes B'_2 + \frac{2}{\pi} \sigma_2 \otimes B'_1 + \sigma_3 \otimes B'_3), \\ \Gamma_2 &= \eta(\sigma_0 \otimes B'_0 - \sigma_1 \otimes B'_1 + \frac{2}{\pi} \sigma_2 \otimes B'_3 + \frac{2}{\pi} \sigma_3 \otimes B'_2), \\ \Gamma_1 &= \eta(\sigma_0 \otimes B''_0 + \frac{2}{\pi} \sigma_1 \otimes B'_2 + \frac{2}{\pi} \sigma_2 \otimes B'_1 - \sigma_3 \otimes B'_3), \\ \Gamma_0 &= \eta(\sigma_0 \otimes B'_0 + \sigma_1 \otimes B'_1 + \frac{2}{\pi} \sigma_2 \otimes B'_3 - \frac{2}{\pi} \sigma_3 \otimes B'_2). \end{aligned} \quad (39)$$

Readers may refer to Appendix A for derivations of the above results.

Every term on the exponent in Eq. (38) is now at least linear in the smallness η or $\|\Omega\|$ and we can invoke the Magnus formula to calculate $\tilde{\Omega}_{\text{PDD}}$. The first order term suggests:

$$\tilde{\Omega}_{\text{DD}}^{(1)} = \sigma_0 \otimes (4B_0 + 2\eta(B'_0 + B''_0)) + \sigma_2 \otimes \frac{4}{\pi} \eta(B'_1 + B'_3). \quad (40)$$

When the gate noise strength η is at least of similar size as the environmental noise ϕ_{SB} —the gate noise is typically stronger than the background noise in many experiments—the first order Magnus term would be of leading order in the Magnus series. And the error phase after PDD is reduced to a linear multiple in η . we upper-bound the size of the first order coupling by:

$$\tilde{\Phi}_{\text{SB}} \approx \left\| \frac{4\eta}{\pi} (B'_1 + B'_3) \right\| \leq \frac{8}{\pi} \eta. \quad (41)$$

This results in the first order breakeven condition:

$$\eta \leq \frac{\pi}{8} \phi_{\text{SB}}. \quad (42)$$

This condition indicates that in order to achieve noise suppression, the gate noise should be bounded by a fraction of the free evolution noise. If any observable improvement from DD is expected, the pulses must be very accurate in the first place. On the other hand if $\eta \ll \phi_{\text{SB}}$, the existence of pulse noise contributes negligibly and the second order term will be similar to the ideal PDD case. Hence to faithfully approximate $\tilde{\Omega}_{\text{DD}}$ in all possible arrangements, we may keep every term that is linear in η and up to second order in the smallness ϕ_{B} or ϕ_{SB} (alternatively: second order in ϕ). Focusing on the interaction part, the result becomes:

$$\tilde{\Omega}_{\text{DD:SB}} = \Omega_{\text{DD:SB}} + \sigma_2 \otimes \frac{4}{\pi} \eta (B_1'' + B_3') + \mathcal{O}(\eta\phi, \eta^2). \quad (43)$$

Bounding the resulting error phase and applying triangular inequality, we have the second order breakeven condition:

$$\eta \leq \frac{\pi}{8} \phi_{\text{SB}} - \frac{\pi}{8} \Phi_{\text{SB}}^{(\text{ub})}. \quad (44)$$

We remark that this is a loose bound and tighter bound can be obtained by performing optimization on $\|\tilde{\Omega}_{\text{DD:SB}}\|$ akin to what we did for the ideal case, rather than using triangular inequality. But the current form is enough to serve our purpose. For verification, we numerically simulated PDD with noisy pulses and compare the results with our break even conditions. The results are shown in Figure 4. To generate the diagrams, for each fixed tuple of parameters $(\phi_{\text{B}}, \phi_{\text{SB}}, \eta)$, we randomly generate 300 sets of the Hermitian matrices $\{\Omega, \Gamma', \Gamma''\}$ over a uniform distribution satisfying the given parameters. Noisy PDD is simulated to obtain the noise reduction ratio $\Phi_{\text{SB}}/\phi_{\text{SB}}$ for each particular set of $\{\Omega, \Gamma', \Gamma''\}$ as inputs. The maximal noise reduction value, which should reflect the worst case performance, is recorded for each fixed $(\phi_{\text{B}}, \phi_{\text{SB}}, \eta)$ parameter. The effects of noisy gates are visualized from two aspects: the shrinking of the noise removal region in the $(\phi_{\text{B}}, \phi_{\text{SB}})$ diagram compared to the ideal case (Figure 4a); and the relation between the maximally tolerable gate noise strength versus the environmental noise strength (Figure 4b). From the figures we can see that our first order condition works well as long as the noise parameter $(\phi_{\text{B}}, \phi_{\text{SB}})$ is weak and the second order condition is applicable even when the noise strength parameters are not necessarily small.

According to Eq. (40), the leading order term that acts non-trivially on the system is proportional to $\eta(B_1'' + B_3')$. As the condition $B_1'' + B_3' = 0$ corresponding to a zero measure set, exact first order decoupling is lost assuming no particular relation between the X and Z noise. This seems to indicate that PDD is impractical as its very usefulness lies in its ability to achieve first decoupling. The counterargument to the above is that the Magnus series is artificial and a non-zero first order term need not indicate large noise. It is more inspirational to relax upon first order decoupling and measuring noise strength in terms of the smallness parameter η . Hence an alternative condition should be considered. The revised first

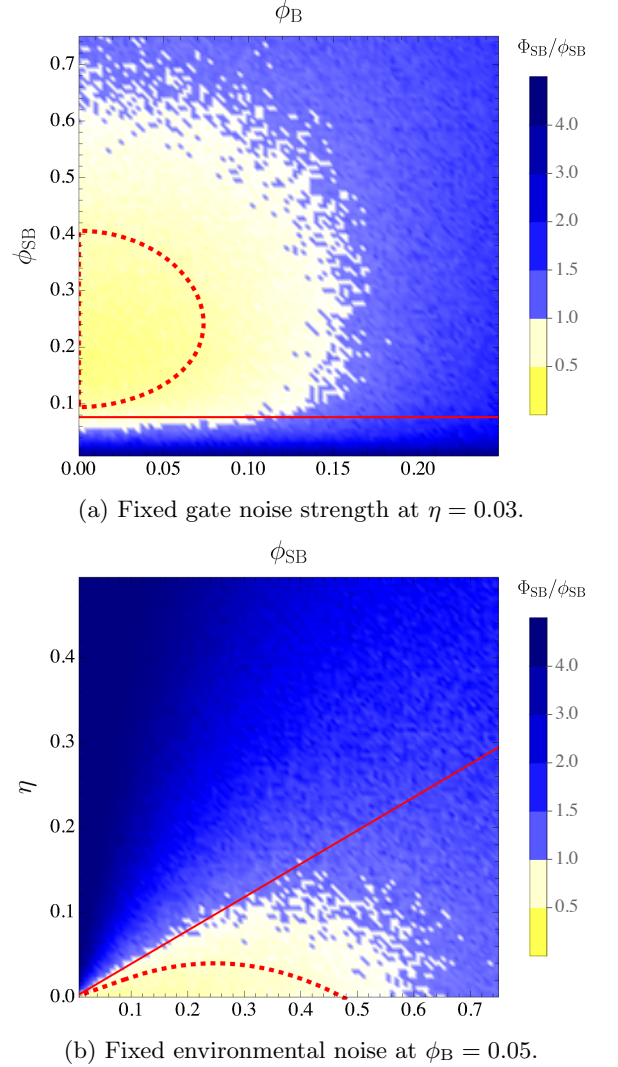


FIG. 4: The noise reduction diagrams for noisy PDD with either fixed gate noise strength or environmental noise strength. The first order and second conditions derived in the main text are indicated with red solid lines and red dashed line respectively for comparison.

order condition demands:

$$\tilde{\Omega}_{\text{DD:SB}}^{(1)} = \mathcal{O}(\eta^2). \quad (45)$$

It can be shown that a sufficient condition for this is the case of “weakly gate-dependent” noise. For the most generic noisy pulses, we can write introduce the gate-noise decomposition $\tilde{P}_i = P_i e^{-i\eta\Gamma_i}$. The weak gate-dependence assumption is demanding that the variance in Γ_i is at most on the order of the smallness parameter η . Hence,

$$\tilde{P}_i = P_i e^{-i\eta\Gamma + \mathcal{O}(\eta^2)}, \quad \forall i. \quad (46)$$

In other words, the dependence of the noise on the gate is of higher order. If this condition is satisfied, then the

time evolution readily reduces to the ideal DD case by replacing $\Omega = \tau H$ for free evolution with an effective Hamiltonian $\tilde{\Omega}$ such that

$$e^{-i\tilde{\Omega}} = e^{-i\eta\Gamma} e^{-i\Omega}. \quad (47)$$

Since Γ and Ω are both small, as required for DD to work, one may invoke the BCH formula to get $\tilde{\Omega}$ as a series of nested commutators. Thus obtained $\tilde{\Omega}$ is accurate to the second order in η . The already established map $\Omega \rightarrow \Omega_{\text{DD}}$ for the idea-pulse DD sequence can then be reused to estimate Ω_{DD} . Naturally, we recover the decoupling behavior for ideal DD. To derive the break-even condition for such weakly gate-dependent noise, we retain the BCH series to the second order term, and then apply triangular inequality to derive the upper bounds for $(\tilde{\phi}_B, \tilde{\phi}_{\text{SB}}) \equiv (\|\tilde{\Omega}_B\|, \|\tilde{\Omega}_{\text{SB}}\|)$:

$$\begin{cases} \tilde{\phi}_B \lesssim (\eta + \phi_B)(1 + \eta + \phi_{\text{SB}}) \equiv \tilde{\phi}_B^{(\text{ub})} \\ \tilde{\phi}_{\text{SB}} \lesssim (\eta + \phi_{\text{SB}})(1 + \eta + \phi_B) \equiv \tilde{\phi}_{\text{SB}}^{(\text{ub})} \end{cases} \quad (48)$$

With this, we can bound the error phase relevant for this noisy-pulse situation under the different schemes by replacing ϕ_B and ϕ_{SB} for the ideal case with their noisy upper bounds:

$$\tilde{\Phi}_{\text{SB}} \leq \Phi_{\text{SB}}(\tilde{\phi}_B^{(\text{ub})}, \tilde{\phi}_{\text{SB}}^{(\text{ub})}) < \phi_{\text{SB}}. \quad (49)$$

Further bounding the right hand side by ϕ_{SB} then gives us the condition for the break-even points.

The above analysis is generic and applies to all DD schemes with gate-independent noise. The resulting bound is general, but could be suboptimal for a specific DD scheme and a specific type of pulse imperfection. Below, we examine a concrete examples of imperfections for PDD: global unitary error in the control gates, which results in the noisy pulse

$$\tilde{P}_i = P_i e^{-i\theta \cdot \sigma}, \quad \forall i. \quad (50)$$

We use the notation $\sigma = (\sigma_1, \sigma_2, \sigma_3)$, and $\theta \equiv (\theta_1, \theta_2, \theta_3)$ with θ_i real constants—taken as small for weak noise—that parameterize the unitary error. This can arise from, for example, a systematic calibration error in the pulse control leading to a consistent over or under rotation. In this situation, $\Gamma \equiv \theta \cdot \sigma$ acts only on the system and does not have a pure bath term. In the appendix, we derive an tighter bound:

$$\begin{cases} \tilde{\phi}_B \leq \phi_B + \frac{1}{3}\theta\phi_{\text{SB}}^2, \\ \tilde{\phi}_{\text{SB}} \leq \phi_{\text{SB}} + \theta^2\phi_{\text{SB}} \end{cases}, \quad (51)$$

where $\theta \equiv \|\theta\|$ is the rotation angle associated with the gate error. Assuming the noise to be small, we may only use the leading order approximation, which is by itself accurate to the second order. Substituting the upper bounds to $\Phi_{\text{SB}}(\tilde{\phi}_B, \tilde{\phi}_{\text{SB}})$ and keeping the leading order

terms, we can predict the following simplified bound for PDD:

$$2\phi_B^2 + \frac{(\theta + \phi_{\text{SB}})^4}{4\phi_{\text{SB}}^2} \leq \frac{1}{16}, \quad (52)$$

From this inequality, the globally maximally allowed θ can be found at $\phi_B = 0$, $\phi_{\text{SB}} = 1/8$ with $\theta_{\text{max}} = 1/8$. Furthermore, we can solve for θ in terms of ϕ_B and ϕ_{SB} to arrive at the noise threshold for the rotation angle:

$$\theta \leq (1 - 32\phi_B^2)^{\frac{1}{4}} \sqrt{\frac{\phi_{\text{SB}}}{2}} - \phi_{\text{SB}} \lesssim \sqrt{\frac{\phi_{\text{SB}}}{2}}. \quad (53)$$

It states that for the regime where the the free evolution noise is small, gate noise should at most be on the order of the square root of the free evolution noise. This is quite different from the generic noise case where the upper bound dependence in ϕ_{SB} is linear.

2. Control signals of finite strength

In reality the control signals have finite strength hence non-zero durations. This deviation from the delta function contributes to an extra source of noise, which adds to the already existing gate control error. In this section, we consider such source of noise.

The simplest example is that of the rectangular signals. Assuming the pulse duration is τ_p , the Hamiltonian for the i th gate interval is split into two constant pieces: $H = H_e + H_c$ for $t_i - \tau_p < t \leq t_i$ under external control, and $H = H_e$ when control is off. The control Hamiltonian is assumed to satisfy $P_i = e^{-i\tau_p H_c}$. Such noise model was considered in the original CDD paper[5]. Here we work out a slightly more involved model, where apart from the finite width, the control Hamiltonian is also infested with a small control error denoted by $\epsilon\Gamma_i$, with $\max(\Gamma_i) = 1, \forall i$ and $\epsilon \ll 1$. Namely, without the environment noise, the i th control gate is $e^{-i(\Omega_{P_i} + \epsilon\Gamma_i)}$, where $e^{-i\Omega_{P_i}}$ is the ideal gate. We write solution to the time-evolution operator as:

$$\tilde{U}(t_i, t_{i-1}) = e^{-i(\Omega_{P_i} + \epsilon\Gamma_i + (\tau_p/\tau)\Omega)} e^{-i(1 - (\tau_p/\tau))\Omega}. \quad (54)$$

Despite being a finite width model, one can convert such case back to the noisy pulse model. To do so, one simply replace the free evolution generator $\Omega \rightarrow (1 - (\tau_p/\tau))\Omega$ and the pulse noise generator $\eta\Gamma$ in Eq. (36) with $\epsilon\Gamma_i + (\tau_p/\tau)\Omega$. In terms of PDD, all we need to obtain the Magnus series is the substitutions $B_i \rightarrow (1 - (\tau_p/\tau))B_i$, $\eta B'_i \rightarrow (\tau_p/\tau)B_i + \epsilon B'_i$ and $\eta B''_i \rightarrow (\tau_p/\tau)B_i + \epsilon B''_i$ in the results for noisy pulses. Using Eq. (40) to derive the first order term, we can estimate the error phase and consequently derivative the breakeven condition as

$$\tilde{\Phi}_{\text{SB}} \lesssim \frac{8}{\pi} \left(\frac{\tau_p}{\tau} \right) \phi_{\text{SB}} + \frac{8}{\pi} \epsilon \leq \phi_{\text{SB}}. \quad (55)$$

The ratio τ_p/τ between the pulse width and the gate interval and is typically considered to be small. Hence

the error phase contribution from the finite width is on the second order smallness if this additional smallness parameter is taken into account. In comparison, the control error directly contributes to the first order smallness. This suggests DD is relatively more tolerant to the finite width error. In the limit where gate control error is negligible, the first order breakeven condition suggests

$$\tau_p/\tau \leq \frac{\pi}{8} \approx 0.39, \quad (56)$$

to be sufficient for noise reduction. This suggests that in order for DD to be useful, the control signal should be more pulse-like. Notice that this condition is derived without any control noise in mind. It puts an upper bound on how fast we are allowed to perform DD even with perfect control. That is, the control signals should be somewhat sparse and allow time for the free evolution instead of jamming together.

Extending beyond the simple rectangular signal model, we now consider a more generic situation where the control signal can be of arbitrarily shape while also accounting for the gate control error. After shifting the time origin for simplicity, the total Hamiltonian within a gate interval can be split into three parts:

$$H(t) = H_g(t) + H_e + H_{ge}(t), \quad 0 \leq t \leq \tau, \quad (57)$$

where $H_g(t)$ is the “gate” Hamiltonian, responsible for the perfect gate; H_e is the system-bath coupling Hamiltonian responsible for the free evolution noise, and $H_{ge}(t)$ is the gate error Hamiltonian responsible for the control error. It is assumed that the control signal $H_g(t)$ is pulse-like and much greater than the two noise terms when control is turned on. We isolate the pure gate part by writing the time evolution as

$$U(\tau) = P_g e^{-i\tilde{\Omega}_g}, \quad (58)$$

where $P_g = U_g(\tau)$ is the gate unitary generated by $H_g(t)$. The interaction picture Hamiltonian defined by $\tilde{H}(t) \equiv U_g^\dagger(t)(H_e + H_{ge}(t))U_g(t)$ completes the time evolution operator $e^{-i\tilde{\Omega}_g}$ at $t = \tau$. Since both $H_e(t)$ and $H_{ge}(t)$ are small, we can approximate $\tilde{\Omega}_g$ by neglecting their higher order commutators:

$$\tilde{\Omega}_g \approx \int_0^\tau U_g^\dagger(t) H_e U_g(t) dt + \int_0^\tau U_g^\dagger(t) H_{ge}(t) U_g(t) dt. \quad (59)$$

We refer the first term in the above as the signal shape term and the second term as the pulse error term.

The signal shape term is very dependent on the control signal $H_g(t)$. But anyway let us formally express the case for X and Z gates as:

$$\begin{aligned} \int_0^\tau U_X^\dagger(t) H_e U_X(t) dt &\equiv \sigma_0 B_0 + \sigma_1 B'_1 + \sigma_2 B'_2 + \sigma_3 B'_3, \\ \int_0^\tau U_Z^\dagger(t) H_e U_Z(t) dt &\equiv \sigma_0 B_0 + \sigma_1 B''_1 + \sigma_2 B''_2 + \sigma_3 B''_3, \end{aligned} \quad (60)$$

where we have note that the B_0 component is invariant under system unitary conjugation. To deal with the time-dependent pulse error term, we simply bound it by

$$\left\| \int_0^\tau U_g^\dagger(t) H_{ge}(t) U_g(t) dt \right\| \leq \int_0^\tau \|H_{ge}(t)\| dt \equiv \eta_{ge}, \quad (61)$$

where the small number η_{ge} absorbs any time-dependence of the control Hamiltonian. Hence we may effectively replace the pulse error term with a gate-dependent error map $\eta_{ge}\Gamma_g$, where Γ_g is Hermitian and $\|\Gamma_g\| \leq 1$. For convenience, let us assume $\Gamma_X = \sum_i \sigma_i \otimes A'_i$ and $\Gamma_Z = \sum_i \sigma_i \otimes A''_i$. After substituting the expressions for the signal shape term and pulse error term into $\tilde{\Omega}_X$ and $\tilde{\Omega}_Z$, we are able to derive the effective Hamiltonian using the Magnus expansion. The first order term yields:

$$\begin{aligned} \tilde{\Omega}_{\text{DD}}^{(1)} &= \sigma_0 \otimes (4B_0 + 2\eta_{ge}(A'_0 + A''_0)) \\ &\quad + \sigma_2 \otimes (2(B'_2 - B''_2) + 2\eta_{ge}(A'_2 - A''_2)). \end{aligned} \quad (62)$$

Hence, the error phase after noisy PDD can be bounded by

$$\tilde{\Phi}_{\text{SB}} \lesssim 2\|B'_2 - B''_2\| + 4\eta_{ge}. \quad (63)$$

For simplicity, let us assume that the gate Hamiltonian responsible for the Pauli gate P_g is proportional to σ_g but the signal strength is arbitrary:

$$H_g(t) = \frac{1}{2}\omega_g(t)\sigma_g, \quad g \in \{X, Z\}, \quad (64)$$

where the time-dependent function $\omega_g(t)$ specifies the shape of the pulse and satisfies the condition that its time integral $\theta_g(\tau) = \pi$, with $\theta_g(t) \equiv \int_0^t \omega_g(t') dt'$. Within $t = 0$ and τ we have $U_g(t) = \cos(\theta_g(t)/2)I - i\sin(\theta_g(t)/2)\sigma_g$. The earlier constraint ensures that $U_g(t)$ recovers the ideal gate at $t = \tau$. After conjugating $H_e\tau = \sum_i \sigma_i \otimes B_i$ with $U_g(t)$, we can write the pulse shape term as:

$$\begin{aligned} \int_0^\tau U_g^\dagger(t) H_e U_g(t) dt &= \sigma_0 B_0 + \\ &\begin{cases} \sigma_1 B_1 + c_X(\sigma_2 B_2 + \sigma_3 B_3) + s_X(\sigma_2 B_3 - \sigma_3 B_2), & g = X; \\ \sigma_3 B_3 + c_Z(\sigma_1 B_1 + \sigma_2 B_2) + s_Z(\sigma_1 B_2 - \sigma_2 B_1), & g = Z, \end{cases} \end{aligned} \quad (65)$$

where the c_g and s_g coefficients are defined as the time averaged trig-functions:

$$c_g \equiv \frac{1}{\tau} \int_0^\tau \cos(\theta_g(t)) dt, \quad s_g \equiv \frac{1}{\tau} \int_0^\tau \sin(\theta_g(t)) dt. \quad (66)$$

For the ideal DD case, which assumes that $H_g(t) = \pi\delta(t - \tau)$ and $H_{ge}(t) = 0$, one can recover $\tilde{\Omega}_X = \tilde{\Omega}_Z = \Omega$ as expected. On the other hand, if we move the pulse position to some $\tau' \in (0, \tau)$, it appears that $\tilde{\Omega}_X \neq \tilde{\Omega}_Z$. But the difference $B'_2 - B''_2$ as defined in Eq. (60) is still zero, which is all that needed for first order decoupling. Focusing on the difference in the B_2 component, we have

$$\begin{aligned} \|B'_2 - B''_2\| &= \|(c_X - c_Z)B_2 + s_X B_3 + s_Z B_1\| \\ &\leq (|c_X - c_Z| + |s_X| + |s_Z|) \cdot \phi_{\text{SB}}. \end{aligned} \quad (67)$$

Here we see that the most general signal shape error is of first order smallness instead of second order for the rectangular signals. We also see that to reduce the signal shape error, one needs to make sure that the control signal should have identical shapes for all gates so that $|c_X - c_Z|$ is negligible; in addition the signal should be sharp to control the size of $|s_g|$ terms.

For concreteness, we use the rectangular signal, but arbitrarily placed in time to illustrate this point. Assuming the shapes for control signals are gate-independent and satisfies

$$\omega_g(t) = \begin{cases} 0, & 0 \leq t < \tau' \\ \pi/\tau_p, & \tau' \leq t < \tau' + \tau_p \\ 0, & \tau' + \tau_p \leq t \leq \tau \end{cases} \quad (68)$$

It is straightforward to calculate that $s_g = \frac{2\tau_p}{\pi\tau}$. Hence the resulting error phase is independent of τ' . After substituting into Eq. (63), we arrive at the breakeven condition

$$\tilde{\Phi}_{\text{SB}} \lesssim \frac{8}{\pi} \left(\frac{\tau_p}{\tau} \right) \phi_{\text{SB}} + 4\eta_{ge} \leq \phi_{\text{SB}}. \quad (69)$$

This produces the same behavior compared to Eq. (55), apart from numerical factors arisen from difference in gate error definitions.

IV. THE ACCURACY THRESHOLD AND DECOUPLING POTENTIAL

We can certainly discuss the break-even point for a particular CDD_n sequence, treating it simply as a “flat-tened” sequence of pulses, and ask when the error phase after CDD_n is reduced:

$$\Phi_{\text{SB},n} < \phi_{\text{SB}}, \quad (70)$$

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

$$\Phi_{\text{SB},n+1} < \Phi_{\text{SB},n} \quad \forall n = 1, 2, \dots \quad (71)$$

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.*

A. Ideal case

Should the accuracy threshold exist, then Eq. (71) 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 CDD_n generator in the ideal case can be estimated as:

$$\begin{aligned} \Omega_n &\approx \sigma_0 \otimes 4^n B_0 \\ &+ \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B_0}^n(B_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}). \end{aligned} \quad (72)$$

It is shown in the appendix B that the above expression indeed captures the leading order behavior of Ω_n . Hence CDD_n indeed achieves n th-order decoupling. Different from the leading order term in the Magnus series, Eq. (72) combines the leading order pure bath part and the interaction part. The pure bath part, whose size can be estimated by $\Phi_{\text{B},n} \simeq 4^n \phi_{\text{B}}$, 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 the error phase:

$$\Phi_{\text{SB},n} \lesssim 2^{n(n+2)} \phi_{\text{B}}^n \phi_{\text{SB}} f_n\left(\frac{\phi_{\text{SB}}}{\phi_{\text{B}}}\right), \quad (73)$$

where the auxiliary f_n function is defined by:

$$f_n(x) \equiv \begin{cases} 1, & x \leq \sqrt{4^n - 1}, \\ \frac{1}{2^n} \sqrt{\left(\frac{4^n - 1}{2x} + \frac{x}{2}\right)^2 + 1}, & x > \sqrt{4^n - 1}. \end{cases} \quad (74)$$

As mentioned in the Preliminaries section, CDD_n achieves n th-order decoupling. This can be understood from the error phase expression: $\Phi_{\text{SB},n}$ after CDD_n is of order ϕ^{n+1} , for $\phi \sim \phi_{\text{B}}, \phi_{\text{SB}}$. That ϕ_{B} enters the error phase should again be of no surprise—as in the PDD case, ϕ_{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 $\Phi_{\text{SB},n}$, an important feature in the CDD. For fixed $\tau_0 \equiv \tau$ as n increases (e.g., if τ 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 ϕ^n factor is eventually overcome by the super-exponentially increasing 2^{n^2} factor. This means there

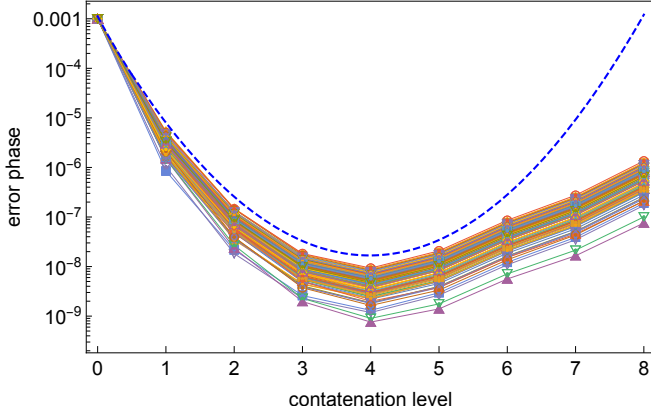


FIG. 5: The error phase as a function of the concatenation level. The thin colored lines are obtained from random samples satisfying $\phi_B = \phi_{SB} = 0.001$. The blue dashed line is our theoretical upper bound for error phase. The maximal concatenation level occurs at 4 in this configuration.

is no accuracy threshold, i.e., there is no level of noise that is weak enough such that $\Phi_{SB,n+1} \leq \Phi_{SB,n}$ 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. 5 plots this situation of fixed τ_0 , for increasing concatenation level n . We observe the initial decrease of $\Phi_{SB,n}$ as n increases, but this turns around eventually.

The upper bound (73) alone is insufficient for predicting when will $\Phi_{SB,n+1} < \Phi_{SB,n}$. 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

$$\begin{aligned} B_{n+1,0} &= 4B_{n,0} \\ B_{n+1,1} &= -4i[B_{n,0}, B_{n,1}] \\ B_{n+1,2} &= -2i[B_{n,0}, B_{n,1}] \\ B_{n+1,3} &= 0 \end{aligned} \quad (75)$$

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

$$n_{\max} \leq -\log_4(\phi_B) - 3/2. \quad (76)$$

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 τ_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. Define the PDD map as \mathcal{D} . The evolution can be thought as

$$\Omega_{n+1} = \mathcal{D}(\Omega_n/4). \quad (77)$$

Rewrite the update rule (75) in terms of the fixed bath operators:

$$\begin{aligned} B_{n+1,0} &= B_{n,0} \\ B_{n+1,1} &= -(1/4)i[B_{n,0}, B_{n,1}] \\ B_{n+1,2} &= -(1/8)i[B_{n,0}, B_{n,1}] \\ B_{n+1,3} &= 0 \end{aligned} \quad (78)$$

This leads to

$$\begin{aligned} \Omega_n &\approx \sigma_0 \otimes B_0 \\ &+ \sigma_1 \otimes (-i)^n 2^{-n^2-n} \text{ad}_{B_0}^n(B_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{-n^2-2n} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}). \end{aligned} \quad (79)$$

Apparently, $\Phi_B = \phi_B$ remains unchanged. The error phase is reduced to

$$\Phi_{SB,n} \lesssim 2^{-n^2} \phi_B^n \phi_{SB} f_n\left(\frac{\phi_{SB}}{\phi_B}\right). \quad (80)$$

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 τ_0 can be in an experiment.

B. 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 $\tilde{\mathcal{D}}$:

$$\tilde{\Omega}_{\text{PDD}} = \tilde{\mathcal{D}}(\Omega). \quad (81)$$

For noisy CDD_n, since we are using the same set of noisy gates as PDD,

$$\tilde{\Omega}_{n+1} = \tilde{\mathcal{D}}(\tilde{\Omega}_n). \quad (82)$$

The threshold condition requires:

$$\Phi_{\text{SB}}(\tilde{\mathcal{D}}(\tilde{\Omega}_n)) \leq \Phi_{\text{SB}}(\tilde{\Omega}_n). \quad (83)$$

This is exactly the breakeven condition for noisy PDD with the free evolution generator Ω replaced by $\tilde{\Omega}_n$. Let us characterize the gate noise level by η , and write the breakeven condition for PDD: $\eta \leq \eta(\phi_B, \phi_{\text{SB}})$. The CDD threshold condition becomes

$$\eta \leq \eta(\tilde{\Phi}_{B,n}, \tilde{\Phi}_{\text{SB},n}). \quad (84)$$

Since the noise level η on the l.h.s. is fixed for a given set of gates, while $\tilde{\Phi}_{B,n}$ and $\tilde{\Phi}_{\text{SB},n}$ 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 $(\tilde{\Phi}_{B,n}, \tilde{\Phi}_{\text{SB},n})$ as points evolving on the graph. Besides the graph itself, we also need to know the update rule $(\tilde{\Phi}_{B,n}, \tilde{\Phi}_{\text{SB},n}) \rightarrow (\tilde{\Phi}_{B,n+1}, \tilde{\Phi}_{\text{SB},n+1})$ for such evolution. This is, by construction of CDD, the same rule for the PDD transformation $(\phi_B, \phi_{\text{SB}}) \rightarrow (\tilde{\Phi}_{B,1}, \tilde{\Phi}_{\text{SB},1})$.

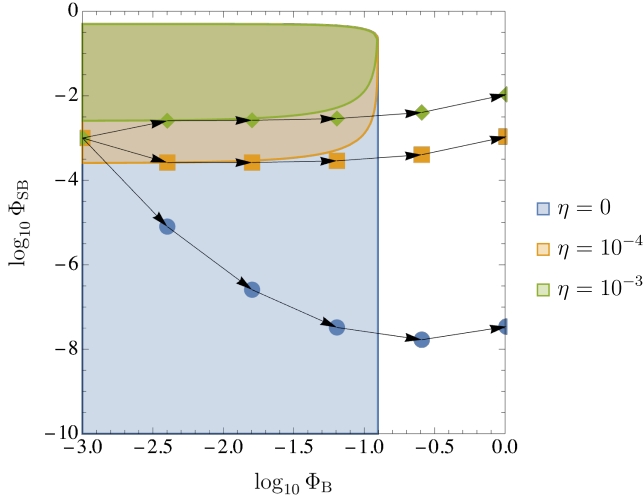


FIG. 6: The evolution of the $(\tilde{\Phi}_B, \tilde{\Phi}_{\text{SB}})$ pair with increasing concatenation level for the ideal and noisy case. The shaped regions are the PDD breakeven region for the noiseless and noisy case respectively.

For noisy CDD, under the computational setting, the size of the pure part part quadruples with every concatenation level:

$$\tilde{\Phi}_{B,n} \approx 4\tilde{\Phi}_{B,n-1} \approx 4^n \phi_B. \quad (85)$$

The evolution of the error phase is similar to that of noisy PDD:

$$\begin{aligned} \tilde{\Phi}_{\text{SB},n} &\lesssim 8\tilde{\Phi}_{B,n-1}\tilde{\Phi}_{\text{SB},n-1} + \frac{8}{\pi}\eta, \\ &\approx (8\phi_B)4^{n-1}\tilde{\Phi}_{\text{SB},n-1} + \frac{8}{\pi}\eta. \end{aligned} \quad (86)$$

Keeping the leading term and the final term,

In Figure 6, we compare the noisy and ideal version of PDD. Due to the presence of noise in the control gates, the maximal concatenation level reduces. The minimal error phase that can be reached by CDD increases as a result.

For the quantum memory setting, the Φ_B part is constant across concatenation levels and the error phase updates according to:

$$\tilde{\Phi}_{\text{SB},n+1} \leq \frac{1}{2}\phi_B\tilde{\Phi}_{\text{SB},n} + \frac{8}{\pi}\eta. \quad (87)$$

Solving the sequence iteration rule leads to the general formula:

$$\tilde{\Phi}_{\text{SB},n} \leq (\phi_B/2)^n \phi_{\text{SB}} + \frac{8\eta/\pi}{1 - \phi_B/2}. \quad (88)$$

Provided that $\phi_B < 2$ and $8\eta/\pi < \phi_{\text{SB}}(1 - \phi_B/2)$, increasing the concatenation level will decrease the noise level. But arbitrary noise suppression is impossible as long as $\eta \neq 0$. The smallest error phase that can be obtained is proportional to the gate noise. We show in Figure 7

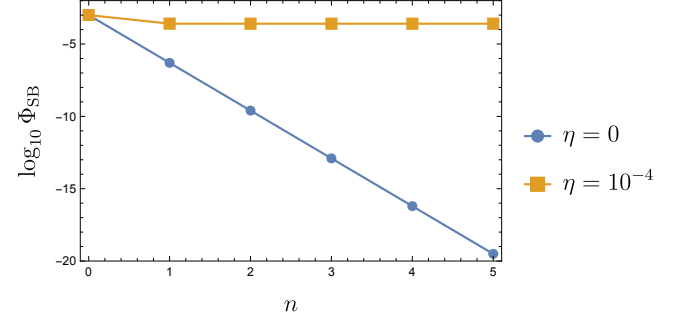


FIG. 7: The evolution of the $(\tilde{\Phi}_B, \tilde{\Phi}_{\text{SB}})$ pair with increasing concatenation level for the ideal and noisy case. The shaped regions are the PDD breakeven region for the noiseless and noisy case respectively.

the evolution of error phase under different levels of concatenation. Starting from a initial $\phi_B = \phi_{\text{SB}} = 10^{-3}$, we see that while in the ideal case, the error phase can be exponentially suppressed, for noisy gates with $\eta = 10^{-4}$, the CDD limit is quickly reached.

V. CROSS MEASURE CONSISTENCY

(Discussions on fidelity, trace distance etc. Compare Xiansong's results here with the theoretical predictions.) To capture more precisely how the error phase relate to the noise strength of a CPTP map, we establish its connection to the channel infidelity. Such connection rely on the assumption that the noise is weak, in the sense that Hermitian operator Ω in Eq. (??) associated with the noise map is small enough.

Invoking the Baker-Hausdorff lemma [16], we expand the output state in Eq. (??) as a series in the powers of Ω . With a simple rearrange,

$$(\mathcal{I} - \mathcal{E})(\rho_S) = - \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \text{tr}_B \text{ad}_{\Omega}^n(\rho_S \otimes \rho_B), \quad (89)$$

where the adjoint operator is defined by the commutator $\text{ad}_{\Omega}(\cdot) \equiv [\Omega, \cdot]$, with ad_{Ω}^n involving n -fold commutation with Ω . According to definition Eq. (??), we can express the channel infidelity as a series as well by taking trace product with ρ_S for each term in Eq. (89),

$$\text{InF}(\mathcal{E}|\rho_S) \equiv \sum_{n=1}^{\infty} \text{InF}^{(n)}(\mathcal{E}|\rho_S). \quad (90)$$

Since the n th-order term $\text{InF}^{(n)}(\mathcal{E}|\rho_S) = \mathcal{O}(\|\Omega\|^n)$, We can examine the leading orders in the infidelity series as an effective approximation.

The first-order infidelity $\text{InF}^{(1)}(\mathcal{E}|\rho_S)$ always vanishes despite the form of Ω , as follows from the equality:

$$\text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega}(\rho_S \otimes \rho_B)) = \text{tr}(\rho_S [\text{tr}_B(\Omega \rho_B), \rho_S]) = 0. \quad (91)$$

To calculate the second-order infidelity, we first note a useful operator identity:

$$\text{tr}_B \text{ad}_{\Omega_B} = 0, \quad \forall \Omega_B \in \mathfrak{h}_B. \quad (92)$$

The proof is straightforward by applying the above operator to a generic linear operator $\sum_{\alpha} S_{\alpha} \otimes B_{\alpha}$ on the combined Hilbert space. We now substitute the Eq. (??) into the double commutator term $\sim \text{ad}_{\Omega}^2$ in Eq. (89) and expand. After applying the Jacobi identity— $\text{ad}_A \text{ad}_B = \text{ad}_B \text{ad}_A + \text{ad}_{[A,B]}$ —and invoking Eq. (91) and Eq. (92), we have

$$\text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega}^2 \rho_S \otimes \rho_B) = \text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega_{SB}}^2 \rho_S \otimes \rho_B). \quad (93)$$

Therefore the second order infidelity $\text{InF}^{(2)}(\mathcal{E}|\rho_S)$ does not involve in Ω_B and is proportional to the square of error phase $\|\Omega_B\|^2$. To bound the infidelity with the error phase, we can evaluate,

$$\begin{aligned} 0 \leq \text{InF}^{(2)}(\mathcal{E}|\rho_S) &= \frac{1}{2} \text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega_{SB}}^2 (\rho_S \otimes \rho_B)) \\ &\leq \frac{1}{2} \|\rho_S \otimes I_B\|_{\infty} \|\text{ad}_{\Omega_{SB}}^2 (\rho_S \otimes \rho_B)\|_1 \\ &\leq 2 \|\Omega_{SB}\|_{\infty}^2 \|\rho_S \otimes \rho_B\|_1 \leq 2 \|\Omega_{SB}\|_{\infty}^2, \end{aligned} \quad (94)$$

where for the upper bound we have applied the Hilbert-Schmidt inequality, the Holder's inequality and the triangular inequality for the Schatten norm class. To show the lower bound, we first introduce the decomposition $\Omega_{SB} = \sum_{i \neq 0} \sigma_i \otimes B_i$ ($\sigma_0 \equiv I_S$) with the generalized Pauli basis for the system, and define the correlation functions $\tilde{B}_{ij} \equiv \text{tr}(\rho_B B_i B_j)$ and $\tilde{S}_{ij} \equiv \text{tr}(\rho_S^2 \sigma_j \sigma_i) - \text{tr}(\rho_S \sigma_i \rho_S \sigma_j)$. Then we have $\text{InF}^{(2)}(\mathcal{E}|\rho_S) = \text{tr}(\tilde{B} \tilde{S})$. It can be easily

verified that both \tilde{B} and \tilde{S} are positive semi-definitive matrices, therefore their trace product is nonnegative.

Furthermore, the full infidelity expression $\text{InF}(\mathcal{E}|\rho_S)$ can be shown to not contain any term linear in the error phase $\|\Omega_{SB}\|$ (see appendix ??). Therefore all higher order terms are indeed smaller than the second order provided $\|\Omega_{SB}\|$ is small, which is true by assumption. In such case, if there is no special symmetry, the second-order infidelity provides a good approximation to the full infidelity expression $\text{InF}(\mathcal{E}) \simeq \|\Omega_{SB}\|_{\infty}^2$, and we may study the second-order infidelity as opposed to the full infidelity, which dramatically simplifies calculations.

On the other hand, the error phase provides only a rough indicator of the channel noise strength and could go wrong if special symmetries present. The bounds in Eq. (94) hints the problem: the error phase is good at measuring the upper bound but fails at the lower bound. In fact, one can think of an extreme example: in some symmetry-protected subspace where $[H, \rho_S \otimes \rho_B] = 0$, the whole evolution is noiseless, thus $\text{InF}(\mathcal{E}|\rho_S) = 0$, but the error phase could be arbitrary. In some other cases where only $\text{InF}^{(2)}(\mathcal{E}|\rho_S) = 0$ and higher order terms persist, the infidelity is no longer approximated by $\|\Omega_{SB}\|^2$.

Inspired from the error phase, we also define two new functions for the channel distance and the channel infidelity by

$$D_{\max}(\alpha, \beta) = \max_{\|\Omega\|=(\alpha, \beta)} D_{\max}(\mathcal{E}), \quad (95)$$

$$\text{InF}_{\max}(\alpha, \beta) = \max_{\|\Omega\|=(\alpha, \beta)} \text{InF}_{\max}(\mathcal{E}), \quad (96)$$

where the $D_{\max}(\mathcal{E})$ and $\text{InF}_{\max}(\mathcal{E})$ are defined in Eq. (??) and Eq. (??) for \mathcal{E} generated by Ω . This allows us to establish a fine-grained hierarchy of functions to gauge performance. For channel distance, $D_{\min}(\alpha, \beta) \leq D_{\min}(\mathcal{E}) \leq D(\mathcal{E}|\rho_S) \leq D_{\max}(\mathcal{E}) \leq D_{\max}(\alpha, \beta)$; and for the channel $\text{InF}_{\min}(\alpha, \beta) \leq \text{InF}_{\min}(\mathcal{E}) \leq \text{InF}(\mathcal{E}|\rho_S) \leq \text{InF}_{\max}(\mathcal{E}) \leq \text{InF}_{\max}(\alpha, \beta)$, where the functions with “min” subscript are defined by replacing the corresponding maximization with minimization. Through the $[H, \rho(0)] = 0$ example we know that $D_{\min}(\alpha, \beta) = \text{InF}_{\min}(\alpha, \beta) = 0$.

VI. CONCLUSION

(To come.)

ACKNOWLEDGMENTS

(Acknowledge Gu Yanwu, Jonas Tan, and Ryan Tiew, who did some of the first calculations of the breakeven points for PDD. Yanwu should be asked if he wants to be a co-author.)

Appendix A: Derivation for equations (39)

Here we explain the detailed steps to derive Eqs. (39). In the following, we make no practical distinction between the notations I, X, Y, Z and $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ except that we prefer to use the former expression for unitary transformations and the latter for Lie algebra elements.

First, using their definitions, we can rewrite $e^{-i\Gamma_i}$ ($i = 0, 1, 2, 3$) in a more uniformed manner:

$$e^{-i\Gamma_3} = \tilde{Z}Z = Ze^{-i\frac{\pi}{2}\sigma_3 - i\eta\sigma_3\Gamma''\sigma_3}, \quad (\text{A1a})$$

$$e^{-i\Gamma_2} = Z\tilde{X}Y = Xe^{+i\frac{\pi}{2}\sigma_1 - i\eta\sigma_2\Gamma'\sigma_2}, \quad (\text{A1b})$$

$$e^{-i\Gamma_1} = Y\tilde{Z}X = Ze^{+i\frac{\pi}{2}\sigma_3 - i\eta\sigma_1\Gamma''\sigma_1}, \quad (\text{A1c})$$

$$e^{-i\Gamma_0} = X\tilde{X} = Xe^{-i\frac{\pi}{2}\sigma_1 - i\eta\Gamma'}. \quad (\text{A1d})$$

Notice that here we have ignored the multiplicative factor $(-i)$ in Eq. (A1b) and Eq. (A1c) since they contribute to global phase factor only.

The goal is to calculate Γ_i to the leading order in the smallness η . To achieve this, we cite a formula from matrix group theory (see e.g. book [16] page 15):

$$\frac{d}{dt}e^X = e^X \frac{1 - e^{-\text{ad}_X}}{\text{ad}_X} \frac{dX}{dt}, \quad (\text{A2})$$

where X is matrix-valued differentiable function of the scalar variable t . With the knowledge for taking derivatives, we can use Taylor expansion to derive the following approximation formula:

$$\begin{aligned} e^{-A}e^{A+tB} &= 1 + t \frac{1 - e^{-\text{ad}_A}}{\text{ad}_A} B + \mathcal{O}(t^2) \\ &= \exp \left[t \frac{1 - e^{-\text{ad}_A}}{\text{ad}_A} B + \mathcal{O}(t^2) \right], \end{aligned} \quad (\text{A3})$$

for general matrix operators A and B independent of the small real number t . Apply Eq. (A3) to Eqs. (A1), we obtain the following expressions for Γ_i :

$$\Gamma_3 = \eta \frac{1 - e^{-\frac{\pi}{2i}\text{ad}_Z}}{\frac{\pi}{2i}\text{ad}_Z} (\sigma_3\Gamma''\sigma_3) + \mathcal{O}(\eta^2), \quad (\text{A4a})$$

$$\Gamma_2 = \eta \frac{1 - e^{-\frac{\pi}{2i}\text{ad}_X}}{-\frac{\pi}{2i}\text{ad}_X} (\sigma_2\Gamma'\sigma_2) + \mathcal{O}(\eta^2), \quad (\text{A4b})$$

$$\Gamma_1 = \eta \frac{1 - e^{-\frac{\pi}{2i}\text{ad}_Z}}{-\frac{\pi}{2i}\text{ad}_Z} (\sigma_1\Gamma''\sigma_1) + \mathcal{O}(\eta^2), \quad (\text{A4c})$$

$$\Gamma_0 = \eta \frac{1 - e^{-\frac{\pi}{2i}\text{ad}_X}}{\frac{\pi}{2i}\text{ad}_X} (\Gamma') + \mathcal{O}(\eta^2). \quad (\text{A4d})$$

The expressions in Eqs. (A4) can be further calculated by writing out the series expansion for the leading order terms in η . Let us focus on Γ_3 , the leading order term of which, according to Eq. (A4a), involves in the operator series:

$$\sum_{k=0}^{\infty} \frac{(-\pi)^k}{(k+1)!} \left(\frac{1}{2i} \text{ad}_Z \right)^k (\sigma_3\Gamma''\sigma_3). \quad (\text{A5})$$

Having introduced the decomposition for Γ'' , we have

$$\sigma_3\Gamma''\sigma_3 = \sigma_0 \otimes B'_0 - \sigma_1 \otimes B'_1 - \sigma_2 \otimes B'_2 + \sigma_3 \otimes B'_3. \quad (\text{A6})$$

We may explicitly calculate the commutator $[Z, \sigma_3\Gamma''\sigma_3]$ and higher order nested commutators in Eq. (A5). By inspection, the results fall in a total of four possible cases:

$$\begin{aligned} \left(\frac{1}{2i} \text{ad}_Z \right)^{4n+1} (\sigma_3\Gamma''\sigma_3) &= -\sigma_2 \otimes B''_1 + \sigma_1 \otimes B''_2, \\ \left(\frac{1}{2i} \text{ad}_Z \right)^{4n+2} (\sigma_3\Gamma''\sigma_3) &= +\sigma_1 \otimes B''_1 + \sigma_2 \otimes B''_2, \\ \left(\frac{1}{2i} \text{ad}_Z \right)^{4n+3} (\sigma_3\Gamma''\sigma_3) &= +\sigma_2 \otimes B''_1 - \sigma_1 \otimes B''_2, \\ \left(\frac{1}{2i} \text{ad}_Z \right)^{4n+4} (\sigma_3\Gamma''\sigma_3) &= -\sigma_1 \otimes B''_1 - \sigma_2 \otimes B''_2, \end{aligned} \quad (\text{A7})$$

for integer $n \geq 0$. Hence, we can calculate Eq. (A4a) by

$$\begin{aligned} (\cdots) &= \sigma_0 \otimes B''_0 + \sigma_3 \otimes B''_3 \\ &\quad - \left(1 - \frac{\pi^2}{3!} + \frac{\pi^4}{5!} + \cdots \right) \sigma_1 \otimes B''_1 \\ &\quad + \left(\frac{\pi}{2!} - \frac{\pi^3}{4!} + \frac{\pi^5}{6!} + \cdots \right) \sigma_2 \otimes B''_1 \\ &\quad - \left(1 - \frac{\pi^2}{3!} + \frac{\pi^4}{5!} + \cdots \right) \sigma_2 \otimes B''_2 \\ &\quad - \left(\frac{\pi}{2!} - \frac{\pi^3}{4!} + \frac{\pi^5}{6!} + \cdots \right) \sigma_1 \otimes B''_2 \\ &= \sigma_0 \otimes B''_0 + \sigma_3 \otimes B''_3 - \frac{\sin \pi}{\pi} (\sigma_1 \otimes B''_1 + \sigma_2 \otimes B''_2) \\ &\quad + \frac{1 - \cos \pi}{\pi} (\sigma_2 \otimes B''_1 - \sigma_1 \otimes B''_2) \\ &= \sigma_0 \otimes B''_0 - \frac{2}{\pi} \sigma_1 \otimes B''_2 + \frac{2}{\pi} \sigma_2 \otimes B''_1 + \sigma_3 \otimes B''_3. \end{aligned} \quad (\text{A8})$$

This leads to the expression for Γ_3 in Eq. (39). One can proceed to verify Γ_2 , Γ_1 and Γ_0 using similar steps.

Appendix B: Analyzing the CDD_n dynamics

The unitary dynamics U_n for 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 $\Omega \equiv \tau H$. At the first level, CDD produces $\Omega_1 \equiv \Omega_{\text{PDD}}$, 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:

$$\Omega_1 = \mathcal{F}(\Omega), \quad \Omega_1^{(m)} = \mathcal{F}^{(m)}(\Omega), \quad (\text{B1})$$

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 through Eq. (13) and Eq. (14). Higher level concatenations are defined by the iteration map

$$\Omega_n = \mathcal{F}(\Omega_{n-1}) = \sum_{m=1}^{\infty} \mathcal{F}^{(m)}(\Omega_{n-1}). \quad (\text{B2})$$

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

$$\Omega_n = \left(\sum_{m=1}^{\infty} \mathcal{F}^{(m)} \right)^n (\Omega) = \sum_{m=1}^{\infty} \Omega_n^{(m)}, \quad (\text{B3})$$

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

The full Ω_n is typically difficult to solve. To estimate Ω_n , it suffice to find an analytically solvable estimator $\hat{\Omega}_n$ 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

$$\Omega_n = \hat{\Omega}_n + \delta\Omega_n. \quad (\text{B4})$$

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 $\hat{\Omega}_n$. This faithfulness condition demands

$$\|\delta\Omega_n\| \equiv \begin{pmatrix} \delta\alpha_n \\ \delta\beta_n \end{pmatrix} \ll \|\hat{\Omega}_n\| \equiv \begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix}, \quad (\text{B5})$$

where the comparison is implied for both components and should be understood by comparing the leading powers of the polynomials. Namely, a “smaller” polynomial should have a larger leading power.

The original CDD paper (Ref.[4]) constructed an estimator by keeping the first two Magnus terms for each iteration step. In our notation:

$$\hat{\Omega}_n \equiv (\mathcal{F}^{(1)} + \mathcal{F}^{(2)})(\hat{\Omega}_{n-1}) = (\mathcal{F}^{(1)} + \mathcal{F}^{(2)})^n(\Omega). \quad (\text{B6})$$

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

$$\begin{aligned} \hat{\Omega}_n &= \sigma_0 \otimes 4^n B_0 \\ &+ \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B_0}^n(B_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}). \end{aligned} \quad (\text{B7})$$

For simplicity, let us make the recognition that $\|B_i\| \simeq \tau$, where \simeq represents that the both sides have the same leading power when written as a polynomial in τ , ignoring any coefficient. According to Eq. (B7), we have,

$$\|\hat{\Omega}_n\| = \begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} \simeq \begin{pmatrix} \tau \\ \tau^{n+1} \end{pmatrix}. \quad (\text{B8})$$

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

$$\begin{pmatrix} \delta\alpha_n \\ \delta\beta_n \end{pmatrix} \lesssim \begin{pmatrix} \tau^3 \\ \tau^{n+2} \end{pmatrix}. \quad (\text{B9})$$

We prove this assertion by mathematical induction.

Proof: At $n = 1$, the estimation error comes solely from the Magnus series truncation for PDD, lead by the third order term:

$$\|\delta\Omega_1\| = \|\sum_{m=3}^{\infty} \Omega_1^{(m)}\| \simeq \|\Omega_1^{(3)}\| \simeq \begin{pmatrix} \tau^3 \\ \tau^3 \end{pmatrix}. \quad (\text{B10})$$

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

$$\begin{aligned} \delta\Omega_{n+1} &= \mathcal{F}(\Omega_n) - (\mathcal{F}^{(1)} + \mathcal{F}^{(2)})(\hat{\Omega}_n) \\ &= \mathcal{F}^{(1)}(\delta\Omega_n) + \sum_{m=2}^{\infty} \mathcal{F}^{(m)}(\Omega_n) - \mathcal{F}^{(2)}(\hat{\Omega}_n), \end{aligned} \quad (\text{B11})$$

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. (B11):

$$\begin{aligned} \|\delta\Omega_{n+1}\| &\lesssim \|\mathcal{F}^{(1)}(\delta\Omega_n)\| + \|\sum_{m=3}^{\infty} \mathcal{F}^{(m)}(\Omega_n)\| \\ &+ \|\mathcal{F}^{(2)}(\Omega_n) - \mathcal{F}^{(2)}(\hat{\Omega}_n)\|. \end{aligned} \quad (\text{B12})$$

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 $\sum_{m=3}^{\infty} \mathcal{F}^{(m)}(\hat{\Omega}_n)$ is small. However Eq. (B11) 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 (B12) 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 the induction hypothesis $\alpha_n \simeq \hat{\alpha}_n \simeq \tau$ and $\beta_n \simeq \hat{\beta}_n \simeq \tau^{n+1}$. The estimation for $\|\mathcal{F}^{(3)}(\Omega_n)\|$ can be done by explicitly calculating the third order Magnus term for PDD; or by making the clever observation that for higher Magnus terms, $B_{i \neq 0}$ is involved in the coupling part for at least once and the pure bath part for at least two times. To bound the third term in (B12), we

need to calculate the difference $\mathcal{F}^{(2)}(\widehat{\Omega} + \delta\Omega) - \mathcal{F}^{(2)}(\widehat{\Omega})$. Using the expression for the second order Magnus term, we obtain

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

Since all three terms are bounded by $(\tau_0^3, \tau_0^{n+3})^\top$, so is their sum. This proves our assertion Eq. (B9). \square

We have now showed that the estimator constructed through Eq. (??)eq:estimator-trunc is indeed faithful. Hence it must be the case that $\widehat{\Omega}_n$ contains at least the leading order terms of the full Ω_n . We are now confident in using the estimator Eq. (B7) to study the full Ω_n of CDD $_n$.

-
- [1] L. Viola, E. Knill, and S. Lloyd, Physical Review Letters **82**, 2417 (1999).
 - [2] L.-M. Duan and G.-C. Guo, Pulse controlled noise suppressed quantum computation (1998), [arXiv:quant-ph/9807072 \[quant-ph\]](#).
 - [3] P. Zanardi, Physics Letters A **258**, 77 (1999).
 - [4] K. Khodjasteh and D. Lidar, Physical review letters **95**, 180501 (2005).
 - [5] K. Khodjasteh and D. A. Lidar, Physical Review A **75**, 062310 (2007).
 - [6] L. Viola and L. F. Santos, Journal of Modern Optics **53**, 2559 (2006).
 - [7] G. S. Uhrig, Physical Review Letters **98**, 100504 (2007).
 - [8] S. Pasini and G. S. Uhrig, Physical Review A **81**, 012309 (2010).
 - [9] Z.-Y. Wang and R.-B. Liu, Physical Review A **83**, 022306 (2011).
 - [10] H. K. Ng, D. A. Lidar, and J. Preskill, Physical Review A **84**, 012305 (2011).
 - [11] W.-J. Kuo and D. A. Lidar, Physical Review A **84**, 042329 (2011).
 - [12] *Quantum Error Correction* (Cambridge University Press, 2013).
 - [13] E. L. Hahn, *Phys. Rev.* **80**, 580 (1950).
 - [14] H. Y. Carr and E. M. Purcell, Physical review **94**, 630 (1954).
 - [15] S. Meiboom and D. Gill, Review of scientific instruments **29**, 688 (1958).
 - [16] W. Rossmann, *Lie groups: an introduction through linear groups*, Vol. 5 (Oxford University Press on Demand, 2006).