

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] ~~[To re-select cited papers. Reduce the number of citations for QEC—QEC is not our main topic, so just cite two or three most seminal works. Also, the original QEC papers are not cited (goes back to Shor and Schumacher or some such early papers in the late 90s...). Perhaps cite a QEC textbook?].~~ 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

(Probably need to re-order the subsections here. To do later.)

A. Basics of DD

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. For a given DD scheme, let L be the number of pulses in the sequence. We denote the i th pulse of the sequence as P_i , and write the sequence as $P_L \dots P_2 P_1$ (I moved the equation here to inline), proceeding from right to left in time. Pulse P_i is applied at time $t_i = t_{i-1} + \tau_i$ where t_{i-1} is the time of the application of P_{i-1} , and t_0 is the start time of the sequence. ~~(Please edit Fig. 1 to illustrate the sentences above, including indicating the time intervals; remove Q_i s from the figure. No need to mention gates here.)~~ We illustrate a typical sequence of DD control gates in Figure 1.

All DD schemes share some fundamental similarities. First, the pulses are chosen from a specified group of transformations $P_i \in \mathbf{G}$, for $i = 1, 2, \dots, L$. Second, any DD sequence satisfies the constraint

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

where 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

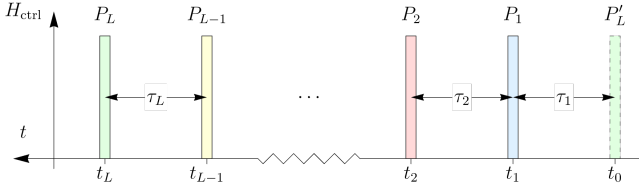


FIG. 1. Illustration of the external control pulses for a typical DD sequence. Different gates are filled with different colors. The dashed gate at t_0 corresponds to the previous DD period.

completion of the sequence. Different strategies could differ in the following aspects: (i) the transformation group 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 $\tau_i \equiv \tau \forall i$, 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 our analysis on the regular-interval schemes of PDD and CDD, differing in sequence length and the specific pulse sequence, but which both employ pulses drawn from the Pauli group.

The physical model we are considering is one where gates, including DD pulses, can be applied at some operating 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. We model the DD pulses as instantaneous gates taking no time. 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 imperfections. 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 de-

grees 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 $\Omega_{DD}^{(1)}$ and $\Omega_{DD}^{(2)}$ terms.) Within the radius of convergence of the Magnus series, $\Omega_{DD}^{(m)}$ decreases in importance as m increases. A DD scheme such that $\Omega_{DD}^{(m)}$ 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 $\|\Omega_{DD}^{(n)}\| \sim \|\tau H\|^n$, compared to $\|\Omega_{DD}^{(1)}\| \sim \|\tau H\|$ without DD.

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

$$H \equiv I \otimes B_I + X \otimes B_X + Y \otimes B_Y + Z \otimes B_Z, \quad (4)$$

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

$$\|\tau H\| \equiv (\|\tau H_B\|, \|\tau H_{SB}\|) \equiv (\phi_B, \phi_{SB}). \quad (5)$$

(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_{DD} of Eqs. eq. (2) and eq. (3), now re-labeled as Ω_{PDD} , can be worked out to be

$$\Omega_{PDD} = \Omega_{PDD}^{(1)} + \Omega_{PDD}^{(2)} + \dots \quad (6)$$

$$\text{with } \Omega_{PDD}^{(1)} = (4\tau) I \otimes B_I \quad (7)$$

$$\Omega_{PDD}^{(2)} = -(2\tau^2) \{ X \otimes 2i[B_I, B_X] + Y \otimes (i[B_I, B_Y] + \{B_X, B_Z\}) \}, \quad (8)$$

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

$$\Omega = (4\tau)H = (4\tau)(I \otimes B_I + H_{\text{SB}}) \quad (9)$$

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 $\|H_{\text{B}}H_{\text{SB}}\|$, not $\|H_{\text{SB}}\|^2$. Also the last sentence of the paragraph following Eq. 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.]

B. Scaling up the protection - concatenated DD

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

$$U_0(t) \equiv e^{-itH_0}, \quad (10)$$

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

$$U_1(\tau_1) \equiv P_L U_0(\tau_0) \dots P_2 U_0(\tau_0) P_1 U_0(\tau_0), \quad (11)$$

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,

$$U_{k+1}(\tau_{k+1}) \equiv P_L U_k(\tau_k) \dots P_2 U_k(\tau_k) P_1 U_k(\tau_k), \quad (12)$$

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

$$U_k(\tau_k) \equiv e^{-i\tau_k H_k} = e^{-i\Omega_k}. \quad (13)$$

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 A re-derives this conclusion with a different and what we believe to be clearer argument than that in the original reference.

C. Quantifying the efficacy of DD

(More to come. To put back parts of the fidelity, etc. discussion, as needed in the remainder of the text.)

1. 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 $U(0, T) \equiv e^{-iT H_{\text{eff}}}$, for some effective time-independent system-bath Hamilton operator H_{eff} . H_{eff} can be split into two pieces: $H_{\text{eff}} \equiv H_{\text{eff},\text{B}} + H_{\text{eff},\text{SB}}$, where $H_{\text{eff},\text{B}} \equiv \frac{I_{\text{S}}}{d_{\text{S}}} \otimes \text{tr}_{\text{S}}(H_{\text{eff}})$ acts on the bath alone, while $H_{\text{eff},\text{SB}} \equiv H_{\text{eff}} - H_{\text{eff},\text{B}}$ contains all the pieces that act nontrivially on the system. $H_{\text{eff},\text{SB}}$ 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,

$$\Phi_{\text{SB}} \equiv T \|H_{\text{eff},\text{SB}}\|. \quad (14)$$

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

$$\Phi \equiv T \|H_{\text{eff}}\| \equiv (\Phi_{\text{B}}, \Phi_{\text{SB}}) \quad (15)$$

(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 $\mathcal{H}_S \otimes \mathcal{H}_B$:

$$\| | (U_S \otimes U_B) \Omega (U_S^\dagger \otimes U_B) | \| = \| | \Omega | \|, \forall U_S(U_B) \in \mathcal{U}(d_{S(B)}), \quad (16)$$

where $\mathcal{U}(d_{S(B)})$ 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.

III. BREAK-EVEN POINT FOR PERIODIC DD

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.

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

$$\text{Memory setting: } \epsilon_{DD} < \epsilon(L), \quad (17)$$

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

$$\text{Computational setting: } \epsilon_{DD} < \epsilon(1), \quad (18)$$

where $\epsilon(1)$ is the noise after bare evolution for a single time step. 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 eq. (18) 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?)

(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.

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,

$$\Phi_{SB} \leq \phi_{SB}. \quad (19)$$

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

$$\Phi_{SB} \simeq \| \Omega_{PDD}^{(2)} \| \leq \phi_{SB}. \quad (20)$$

Using the expression for $\Omega_{PDD}^{(2)}$ from Eq. eq. (8), we have

$$\begin{aligned} \| \Omega_{PDD}^{(2)} \| &\leq 2\tau^2 (2\| [B_I, B_x] \| + \| [B_I, B_Y] \| + \| \{B_X, B_Z\} \|) \\ &\leq 4\phi_B \tau (2\| B_X \| + \| B_Y \|) + 4\tau^2 \| B_X \| \| B_Z \| \\ &\leq 12\phi_B \phi_{SB} + 4\phi_{SB}^2 \end{aligned} \quad (21)$$

noting that $\phi_B \equiv \tau\|H_B\| = \tau\|B_I\|$, and that $\phi_{SB} \equiv \tau\|H_{SB}\| = \tau\left[\sum_{i=X,Y,Z}\|B_i\|^2\right]^{1/2} \geq \tau\|B_i\|$ 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 eq. (20), it then suffices to require

$$12\phi_B + 4\phi_{SB} \leq 1. \quad (22)$$

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. 2? 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, which was explained in the error phase section in version 1, is that the norm need to satisfy $\|I\| = 1$ for whatever dimensionality of the identity. —Jiaan)

$$\left\|\sum_{i=1}^3 \sigma_i \otimes B_i\right\| \approx \sqrt{\sum_{i=1}^3 \|B_i\|^2}. \quad (23)$$

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

$$\begin{aligned} \beta_{PDD}^2 &\approx \|-4i[B_0, B_1]\|^2 + \|-2i[B_0, B_2] - 2\{B_1, B_3\}\|^2 \\ &\leq (8\alpha\|B_1\|)^2 + (4\alpha\|B_2\| + 4\|B_1\|\|B_3\|)^2, \end{aligned} \quad (24)$$

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

$$\max_{x_1^2+x_2^2+x_3^2=1} (2\alpha x_1)^2 + (\alpha x_2 + \beta x_1 x_3)^2 \leq \frac{1}{16}. \quad (25)$$

Solution of this constrained optimization problem leads to:

$$\left(\frac{\beta}{\sqrt{3}} \leq \alpha \leq \frac{1}{8}\right) \text{ or } \left(10\alpha^2 + \frac{9\alpha^4}{\beta^2} + \beta^2 \leq \frac{1}{4}\right). \quad (26)$$

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

The criteria (??) and (26) 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:

$$\alpha + \beta \leq \frac{1}{4}. \quad (27)$$

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 2, 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 (26), together with the absolute convergence region (27) 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” (26), 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

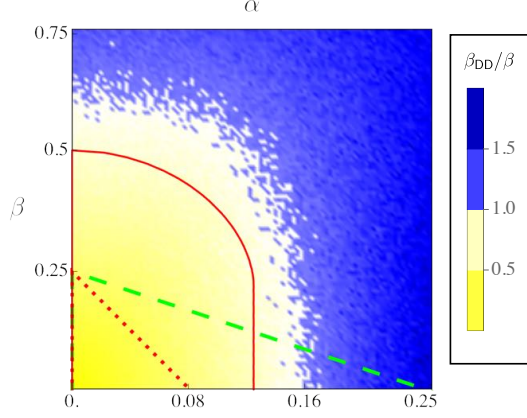


FIG. 2. The noise reduction ratio β_{DD}/β obtained through maximization over random bath operators plotted in the (α, β) parameter plane. In comparison, the theoretical boundaries for noise reduction are also plotted, with dotted red line for (??) and solid red line for (26), also with the boundary for the absolute-convergence condition (27) in dashed green line. Note the ranges for α and β are chosen differently for better visual effect.

line), where our derivation should be unreliable in principle. An explanation is that the absolute-convergence criteria (27), 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. But for safety, the absolute convergence region will also be indicated.

B. Noisy pulses

Above, we treated the DD pulses in an ideal manner, with P_i s as instantaneous, noise-free gates. In practice, of course, P_i s are noisy. Without loss of generality, we can write a noisy pulse \tilde{P}_i , now regarded as a map on states, not just a unitary operator, as $\tilde{P}_i \equiv P_i \circ \mathcal{E}_i$ with $\mathcal{E}_i \equiv \mathcal{P}^{-1} \circ \tilde{P}_i$, and $\mathcal{P}_i(\cdot) \equiv P_i(\cdot)P_i^\dagger$. \mathcal{E}_i describes the noise associated with the i th pulse. For the finite-width pulses analyzed in past DD papers (cite), \tilde{P}_i simply describes the joint evolution of the system and bath according to the Hamilton operator $H + H_{DD,i}$, where H is the bare system-bath Hamilton operator (i.e., no DD) while $H_{DD,i}$ generates the DD pulse P_i , so that H leads to the pulse imperfections in the course of its application. Here, we allow for more general noise, including control noise that does not originate from H . \tilde{P}_i —and consequently \mathcal{E}_i —can be taken as a completely positive (CP) and trace-preserving (TP) map on the system and the bath, a description that captures most cases in practice.

If \mathcal{E}_i is a unitary map, then $\mathcal{E}_i(\cdot) = V_i(\cdot)V_i^\dagger$ with V_i a unitary operator that we write as $V_i \equiv e^{-i\tau\Gamma_i}$, where

Γ_i is a Hermitian operator to be viewed as the effective Hamilton operator generating the noise in the i th pulse over the time interval τ . Even for a non-unitary CPTP map \mathcal{E}_i , we can dilate to include an ancillary system such that \mathcal{E}_i can be viewed as a unitary map on the system-bath-ancilla composite, with \mathcal{E}_i arising after tracing over the ancilla. For simplicity of the analytical treatment below, we will always assume this dilation, and treat all noise as unitary \mathcal{E}_i with an effective Hamilton operator Γ_i . The bath is then assumed to include the ancillary system needed for the dilation. With the noisy pulses, the evolution of the system and bath under DD thus becomes

$$\begin{aligned} \tilde{U}_{DD} &\equiv P_K e^{-i\tau\Gamma_K} e^{-i\tau H} P_{K-1} \dots \\ &\dots P_2 e^{-i\tau\Gamma_2} e^{-i\tau H} P_1 e^{-i\tau\Gamma_1} e^{-i\tau H} \equiv e^{-i\tilde{\Omega}_{DD}}. \end{aligned} \quad (28)$$

Γ can be taken to have no pure-bath term, i.e., it does not have a term proportional to I on the system. Such a term will give rise only to an overall phase on the system and cannot result in observable imperfections in the pulses. (To be merged with new text.) From its definition, we know that an extension of an irrelevant Hilbert space does not change the error phase. So the change in the error phase results from the noise in the gates rather than the choice of ancilla space.

We can examine the simplest case where the noise is gate-independent, i.e., $\Gamma_i = \Gamma \forall i$. The problem reduces to the ideal DD case by replacing H in our earlier analysis by \tilde{H} such that

$$e^{-i\tau\tilde{H}} = e^{-i\tau\Gamma} e^{-i\tau H}. \quad (29)$$

This calls immediately for the BCH formula for getting \tilde{H} in terms of Γ and H ,

$$\tilde{H} = \Gamma + H - \frac{i}{2}[\Gamma, H] - \frac{1}{12}([\Gamma, [\Gamma, H]] - [H, [\Gamma, H]]) + \dots, \quad (30)$$

where higher-order terms are nested commutators of Γ and H . Assuming Γ and H small, as required for DD to work, we can truncate the series and write $\tilde{H} = \tilde{H}_B + \tilde{H}_{SB} \simeq \Gamma + H$ with $\tilde{H}_B \simeq H_B$ the bath-only term, and $\tilde{H}_{SB} \simeq \Gamma + H_{SB}$.

With this, we can estimate the error phase relevant for this noisy-pulse situation under the PDD scheme. To do that, we simply need to replace $\phi_{SB} \equiv \tau\|H_{SB}\|$ in Eq. eq. (21) by $\tau\|\Gamma + H_{SB}\| \leq \eta + \phi_{SB}$, for $\eta \equiv \tau\|\Gamma\|$, giving us the estimate $\Phi_{SB} \lesssim 12\phi_B(\eta + \phi_{SB}) + 4(\eta + \phi_{SB})^2$. For PDD to work, then, it suffices to require

$$12\phi_B(\eta + \phi_{SB}) + 4(\eta + \phi_{SB})^2 \leq \phi_{SB}. \quad (31)$$

This gives us the condition for the break-even point, a joint requirement on ϕ_B , ϕ_{SB} , and η such that PDD improves on the no DD case.

[The original analysis by Jiaan (?)—see previous version of the draft—assumes that Γ acts only on the system + ancilla, not on the bath. This is actually not a good assumption. In general, Γ acts on all three. Having

Γ not act on the bath is appropriate for purely gate-control noise, but excludes then the older DD analyses that treats finite-width pulses with imperfections arising from the same joint evolution of the system with the bath from H_{SB} . Both control noise and finite width of the pulses are important practical realities. We can insist on having Γ describe only control noise and hence act only on system + ancilla, but then we should also put in the finite-width analysis on top of that, which is just as unavoidable as control noise, but that's not done right now in Jiaan's analysis.

See if my simple analysis above suffices, without having to assume that Γ acts only on system + ancilla, at least for the general bound before we specialize to a particular DD scheme. The \tilde{H}_{SB} result gives the same as Jiaan's $\phi_{\text{SB}}(\equiv \beta)$ expression. But because I am taking only the lowest order term, I only have $\tilde{H}_{\text{B}} \simeq H_{\text{B}}$. **Please see if you can add in η corrections to it without resorting to requiring Γ act only on the system + ancilla and not the bath. It's probably not important if we can't, but it's worth at least a bit of thought.]**

The above analysis is conceptually simple and applies to all DD schemes. The resulting bound is general, but could be suboptimal for a specific DD scheme and a specific type of pulse imperfection. Below, we examine **three (?) concrete examples of imperfections for PDD. (Suggestion: Add a third example where we discuss unitary control noise together with finite-width pulses. When the unitary control noise vanishes, we should recover past known results having to do with finite-width pulses. Finite-width pulses are modeled easily (see past DD analyses): Rather than writing P_i as a single unitary gate, we write $P_i = e^{-i\tau_0 H_i}$, where τ_0 is the pulse duration—the finite width—and H_i is the driving Hamilton operator that generates the pulse P_i . This H_i then is added to the total system-bath Hamilton operator during the pulse duration.)**

1. Unitary control noise

Consider a unitary error in the control of the gate, resulting in a noisy pulse

$$\tilde{P}_i = P_i e^{-i\tau \vec{\theta} \cdot \vec{\sigma}}. \quad (32)$$

where $\vec{\sigma} \equiv (X, Y, Z) = (\sigma_1, \sigma_2, \sigma_3)$, and $\vec{\theta} \equiv (\theta_1, \theta_2, \theta_3)$ with θ_i real constants—taken as small for weak noise—that parameterize the unitary error. For simplicity, we assume gate-independent noise, so the same $\vec{\theta}$ applies to all pulses. In this situation, $\Gamma \equiv \vec{\theta} \cdot \vec{\sigma}$ acts only on the system. This can arise from, for example, a systematic calibration error in the pulse control leading to a consistent over or under rotation. One could also include in this model a random over/under-rotation by drawing θ from a specified distribution that describes the randomness in the control. Here, we restrict to the systematic error case and note only that our analysis applies to this

random error situation simply by doing an average over the distribution of $\vec{\theta}$.

We can write the effective noisy Hamilton operator as

$$\tilde{H} \equiv \vec{\theta} \cdot \vec{\sigma} + H', \quad (33)$$

where $H' \equiv \sum_{i=1,2,3} \sigma_i \otimes B'_i$ s, with B'_i to be related to the B_i s in the original H . Keeping up to the second-order term in the BCH series of Eq. (30), straightforward algebra yields $B'_0 \simeq B_0$, and $B'_i \simeq B_i + \sum_{jk=1,2,3} \epsilon_{ijk} \theta_j B_k$, for $i = 1, 2, 3$, and ϵ_{ijk} is the completely anti-symmetric tensor. To lowest order in $\vec{\theta}$, we can regard (B'_1, B'_2, B'_3) as a rotation of (B_1, B_2, B_3) with respect to the $\vec{\theta}$ -axis. **(To check this and merge with the new text.)** On the other hand, we know from its definition that the error phase is the invariant with respect to a basis change of the system, which acts as a 3D rotation on the vector (B_1, B_2, B_3) . As a result, the error phase of Ω' remains identical to β up to the second order in θ .

To conclude, the two-component norm of Ω' can be estimate by

$$\|\Omega'\| = \begin{pmatrix} \tilde{\alpha} \\ \beta' \end{pmatrix} \leq \begin{pmatrix} \alpha + \frac{1}{3}\theta\beta^2 \\ \beta + \theta^2\beta \end{pmatrix}. \quad (34)$$

Assuming the noise to be small, we may only use the leading order approximation, knowing that it is accurate to the second order. **Where do the $\tilde{\alpha}$ and β' expressions come from? I see that the $\tilde{\alpha}$ one comes from setting $\Gamma_{\text{SA}} = \tau \vec{\theta} \cdot \vec{\sigma}$ and hence $\eta = \tau \|\Gamma_{\text{SA}}\| \propto \theta$, and then plugging that into Eq. (69) of the earlier draft, with $\eta_0 = 0$. But why is the proportionality constant 1? What is the norm used here? Same question for β' : Where does the coefficient for the $\theta^2\beta$ come from? I'm understanding that we don't have a linear-in- θ term because of the rotation argument above, but that does not explain the coefficient of 1 for the quadratic term. Is it just a matter of algebra using the BCH formula for the next-order terms?**

Also, how is θ related to the vector $\vec{\theta}$? What norm is being used for this?

(I'm holding off on editing the rest of this section until I get better clarity on the above questions. Retain only the PDD discussion here; CDD discussion goes into the next section. Also need to put back the argument leading up to Eq. (69) of the earlier draft for when Γ acts purely on the system and not on the bath.)

2. Unitary control noise with finite-width pulses

(To add?)

3. Gate-dependent noise

(To consider whether we want to combine the gate-indep and gate-dep noise sections. Treat gate-indep noise

case as a special case of gate-dep noise.) Lastly, we consider gate-dependent noise, where \mathcal{E}_i can depend on i . Rather than having \mathcal{E}_i to be different for every i , we assume—as is typical in practical situations—that the pulse noise depends on what gate the pulse corresponds to, not where it occurs in the DD sequence. For PDD, the sequence comprises only X and Z gates. This amounts to writing the noisy X and Z pulses as

$$\tilde{X} = X e^{-i\tau\Gamma_X} \quad \text{and} \quad \tilde{Z} = Z e^{-i\tau\Gamma_Z}. \quad (35)$$

Here, Γ_X and Γ_Z can generally act on the system and bath (including possibly an ancillary system). Both are assumed to have no pure-bath only term, as any such term adds only a global phase on the system and hence of no consequence. We define

$$e^{-i\tau\tilde{H}_k} \equiv e^{-i\tau\Gamma_k} e^{-i\tau H}, \quad (36)$$

for $k = X$ and Z . The time evolution over the noisy PDD sequence then becomes

$$\begin{aligned} e^{-i\tilde{\Omega}_{\text{PDD}}} &= Z e^{-i\tau\tilde{H}_Z} X e^{-i\tau\tilde{H}_X} Z e^{-i\tau\tilde{H}_Z} X e^{-i\tau\tilde{H}_X} \\ &= e^{-i\tau Z \tilde{H}_Z Z} e^{-i\tau Y \tilde{H}_X Y} e^{-i\tau X \tilde{H}_Z X} e^{-i\tau \tilde{H}_X}. \end{aligned} \quad (37)$$

We can now apply the Magnus formula Eq. (??) (fix the eq. reference) to evaluate $\tilde{\Omega}_{\text{PDD}}$. Writing $\tilde{H}_X \equiv \sum_i \sigma_i \otimes B'_i$ and $\tilde{H}_Z \equiv \sum_i \sigma_i \otimes B''_i$, the first two Magnus terms are, after some straightforward algebra,

$$\begin{aligned} \tilde{\Omega}_{\text{PDD}}^{(1)} &= 2\tau [\sigma_0 \otimes (B'_0 + B''_0) + \sigma_2 \otimes (B'_2 - B''_2)] \\ \tilde{\Omega}_{\text{PDD}}^{(2)} &= \tau^2 \left[\sigma_0 \otimes i([B'_0, B''_0] + [B'_1, B''_1] - [B'_2, B''_2] - [B'_3, B''_3]) \right. \\ &\quad + \sigma_1 \otimes (-i[B'_0 + B''_0, B'_1 + B''_1] + \{B'_2 - B''_2, B'_3 - B''_3\}) \\ &\quad + \sigma_2 \otimes (-i[B'_0, B'_2] - i[B''_0, B''_2] - \{B'_1, B'_3\} - \{B''_1, B''_3\}) \\ &\quad \left. + \sigma_3 \otimes (i[B'_0 + B''_0, B'_3 - B''_3] + \{B'_1 + B''_1, B'_2 - B''_2\}) \right]. \end{aligned} \quad (38)$$

When the noise is gate-independent, i.e., when $\Gamma_X = \Gamma_Z$ and hence $\tilde{H}_X = \tilde{H}_Z$, these two expressions reduce back to the ideal PDD situation, consistent with our earlier discussion in Sec. IIIB1 that the gate-independent imperfections appear only in the third-order Magnus term. (Check for consistency with previous section—this should reduce back to the gate-independent noise case. Should make a remark in the gate-indep noise section that the Pauli averaging also averages away the gate noise to first order. Why is the second-order term also the same as the ideal case? I have checked that it is so, but it is right now somewhat unclear to me; to revisit and explain this in the earlier section.)

With gate-dependent noise, however, the coupling part of the first-order Magnus term no longer vanishes. (Note that this means that this is a genuinely new effect, different from finite-width analyses from earlier papers – the finite-width noise is all gate-independent, since the noise

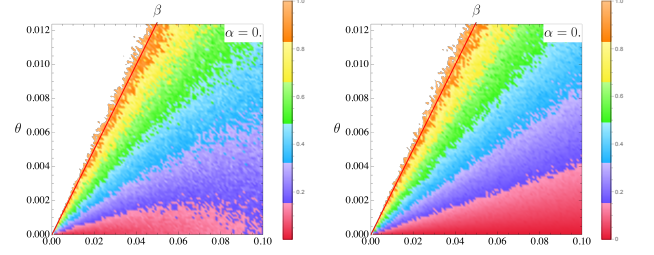


FIG. 3. (Which one is for PDD? Why do some points violate the threshold condition? Is it because of numerical—sampling?—inaccuracies?) The noise threshold compared for PDD and CDD₃ simulated at $\alpha = 0$. The red solid line represents the noise threshold prediction $\eta \leq \beta/4$. different colors represents different levels of noise reduction ratio β_{DD}/β after numerically maximized over random bath operator instances.

is generated by H itself. And this is more significant, since it appears already at first order. To emphasize this point in the text.) The term that acts nontrivially on the system in $\tilde{\Omega}_{\text{PDD}}^{(1)}$ is proportional to $B'_2 - B''_2$. From Eq. eq. (36) and the BCH formula, we can estimate

$$\tilde{H}_X - \tilde{H}_Z \simeq (\Gamma_X - \Gamma_Z) - \frac{i}{2}[\Gamma_X - \Gamma_Z, H]. \quad (39)$$

When the gate noise is at least of similar strength as the noise from H_{SB} —the gate noise is in fact stronger than the background noise in many experiments—this difference is dominated by the first term $\Gamma_X - \Gamma_Z$. Analogous to the gate-independent noise situation, we take $\eta \equiv \tau \max\{\|\Gamma_X\|, \|\Gamma_Z\|\}$. Assuming no particular relation between the X and Z noise, we expect $\tau\|\Gamma_X - \Gamma_Z\| \sim 2\eta$, which translates into an estimate of $\tau\|B'_2 - B''_2\| \sim 2\eta$. In fact, we can upper bound $\tau\|B'_2 - B''_2\| \leq 2\eta$, so that the error phase for PDD can be estimated as $\Phi_{\text{SB}} \simeq 2\tau\|B'_2 - B''_2\| \leq 4\eta$.

With this estimate, the condition for the break-even point for PDD under such gate-dependent noise can be written as

$$4\eta \leq \phi_{\text{SB}}. \quad (40)$$

(Please provide more details for the figure, on how the numerics are done. For example, what random bath operators are we referring to? The B 's and B'' 's? What do you mean by “random” here? Uniform distribution over what? Initial system and bath states? etc.) In Figure 3, we simulate both PDD and CDD₃ with the gate-dependent noise model. We find that our prediction works pretty well.

C. Different measures

(To come.) (Do I understand correctly that all the numerics for this are for PDD only?)

IV. ACCURACY THRESHOLD FOR CONCATENATED DD

(Generalization of the previous PDD section; CDD content from Secs. IVA-D of draft by Jiaan & Xiansong. The main goal here is to show that an accuracy threshold for DD, under the CDD scheme for scaling up, fails to exist. This means that there is a maximum level for CDD, beyond which further levels only make things worse. To discuss the differences from using different measures.)

Next, we turn to the situation of CDD. We can certainly discuss the break-even point for a particular CDD_n sequence, treating it simply as a “flattened” sequence of pulses, and ask when the error phase after CDD_n is smaller than ϕ_{SB} as was done for PDD in the previous section. (I realised that the CDD threshold analysis in the earlier draft is actually all about this only. We can consider adding some part of the content that we already have to the previous section. However, we still need to add this accuracy threshold analysis as described below – to me, this accuracy threshold is the interesting angle from the perspective of seeing CDD as a procedure for scaling up the basic DD scheme. This exactly mirrors the break-even point versus accuracy threshold discussion in quantum error correction and fault tolerance theory (which offers a way to scale up to better protection).)

However, what is more interesting is the question of the so-called accuracy threshold for CDD. As described in the Preliminaries section, CDD tries to achieve better and better noise removal by concatenating the same basic DD scheme—PDD for our discussion here—to higher and higher levels. Here, we are interested in understanding how the noise removal improves as the DD scheme scales up. 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 parameters of the problem such that increasing the CDD level always leads to improved noise removal capabilities. An accuracy threshold for CDD exists if the parameters governing the noise of the situation satisfying a condition (usually one where the noise is weak enough) such that

$$\Phi_{\text{SB},n+1} \leq \Phi_{\text{SB},n} \quad \forall n = 1, 2, \dots \quad (41)$$

Below, we examine this for both the ideal CDD case where pulses are perfect, and the noisy CDD case where imperfections in the pulses are allowed.

A. Ideal case

We first consider CDD with perfect pulses. A detailed analysis (see App. A) yields the error phases for CDD_n

as

$$\begin{aligned} \Phi_{\text{B},n} &\simeq 4^n \phi_{\text{B}}(\tau_0), \\ \Phi_{\text{SB},n} &\lesssim 2^{n(n+1)} \phi_{\text{B}}(\tau_0)^{n-1} \phi_{\text{SB}}(\tau_0) (2^n \phi_{\text{B}} + \phi_{\text{B}} + \phi_{\text{SB}}), \end{aligned} \quad (42)$$

where, to remind the reader, $\phi_{\text{B}} \equiv \phi_{\text{B}}(\tau_0) \equiv \tau_0 \|H_{\text{B}}\|$, and $\phi_{\text{SB}} \equiv \phi_{\text{SB}}(\tau_0) \equiv \tau_0 \|H_{\text{SB}}\|$, noting the dependence of both quantities on τ_0 , the time interval between two consecutive pulses in the CDD sequence. (To reconsider: We need both an upper and lower bound for $\Phi_{\text{SB},n}$, which we have from the estimator [Eq. eq. (A5)]. For use in Eq. eq. (41), we should have the lower bound on the RHS of that condition, and upper bound expression on the LHS, for a sufficient condition.) 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 factors of 2^n and 2^{n^2} in $\Phi_{\text{SB},n}$, a new feature in the CDD case (Do these factors also appear in the Khodjasteh-Lidar analysis? Please check and put in a sentence about this.). 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 is no accuracy threshold, i.e., there is no level of noise that is weak enough such that $\Phi_{\text{SB},n+1} \leq \Phi_{\text{SB},n}$ 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. 4 plots this situation of fixed τ_0 , for increasing concatenation level *n*. We observe the initial decrease of $\Phi_{\text{SB},n}$ as *n* increases, but this turns around eventually.

(To reconsider with my lower/upper bound comment above.) For higher concatenation level to make sense, increasing concatenation level should decrease the error phase. To derive a sufficient condition for the maximal concatenation level, we compare the error phase upper bound in Eq. (A13) of the CDD_n to that of the CDD_{n+1},

$$\frac{2^{(n+1)(n+2)} \alpha^n \beta (2^{n+1} \alpha + \alpha + \beta)}{2^{n(n+1)} \alpha^{n-1} \beta (2^n \alpha + \alpha + \beta)} \leq 4^{n+1} 2\alpha.$$

The maximal level is reached when there is a contraction. By requiring the r.h.s. to be smaller than 1, we obtain the bound for the maximal concatenation level:

$$n_{\text{max}} \leq -\log_4 \alpha - \frac{3}{2}. \quad (43)$$

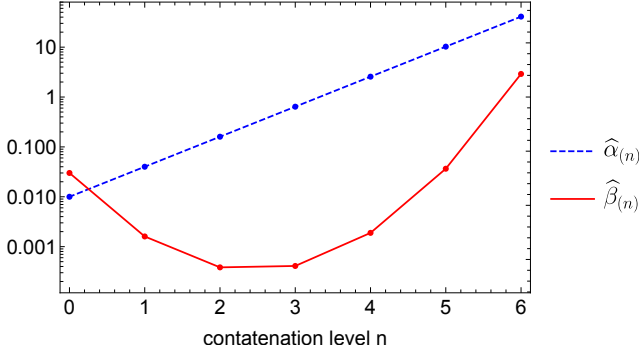


FIG. 4. (To merge with new text; no mention of estimators in current new text.) The estimator size for $\hat{\alpha}_n$ and the error phase $\hat{\beta}_n$ as a function of concatenation level n . In the graph, the bare Hamiltonian is set to take on the value $\alpha = 0.01$ and $\beta = 0.01$.

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. This requires, of course, the ability to do faster and faster pulse switching.

In this case, it is more illuminating to make use of $\tau_0 \|H_B\|$ and $\tau_0 \|H_{SB}\|$, in place of ϕ_B and ϕ_{SB} , in our estimate of the CDD error phases, so that that the n dependence in τ_0 is explicit. (I suggest we analyze this case as well, to see if an accuracy threshold exists. Ji-aaan - could you take a quick look at this?) 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

(To add.)

V. 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: Analyzing CDD_n

(To be edited.) It was suggested by the authors of CDD that the decoupling order for CDD_n is n [4]. This statement is correct, but their proof is logically flawed. Given its importance, here we explain why the original proof is insufficient and provide a rigorous proof for the decoupling order.

For the generic CDD_n, it is infeasible to flatten out the gates and calculate the combined Magnus series directly. Instead it is easier to study the iteration map

$$\Omega_{(n)} = \Omega_{\text{PDD}}(\Omega_{(n-1)}) = \sum_{m=1}^{\infty} \Omega_{\text{PDD}}^{(m)}(\Omega_{(n-1)}), \quad (\text{A1})$$

where the update map Ω_{PDD} is defined in Eq. (??). Backtracking the map from level n to level 0, we have $\Omega_{(n)} = \Omega_{\text{PDD}} \circ \dots \circ \Omega_{\text{PDD}}(\Omega) \equiv (\Omega_{\text{PDD}})^n(\Omega)$. In the end, the effective Hamiltonian $\Omega_{(n)}$ can be expanded as:

$$\Omega_{(n)} = \sum_{m=1}^{\infty} \Omega_{(n)}^{(m)}(\Omega), \quad (\text{A2})$$

where $\Omega_{(n)}^{(m)}$, with the size of $\mathcal{O}(\|\Omega\|^m)$, is the m th order Magnus term in Ω . Notice that this series is different from the series in Eq. (A1), where each $\Omega_{\text{PDD}}^{(m)}$ generate the m th order Magnus term in $\Omega_{(k-1)}$.

To calculate the leading order Magnus term in Ω , we do not directly solve $\Omega_{(n)}$ but instead constructing an estimator $\hat{\Omega}_{(n)}$ which is easier to solve. Formally we have an estimator-error pair $(\hat{\Omega}_{(n)}, \delta\Omega_{(n)})$ such that

$$\Omega_{(n)} = \hat{\Omega}_{(n)} + \delta\Omega_{(n)}. \quad (\text{A3})$$

The estimator $\hat{\Omega}_{(n)}$ is faithful if it captures the leading order behavior of $\Omega_{(n)}$. This is defined separately for the pure-bath part and the coupling part, as they are of different orders. In other words, a faithful estimator

should contain the first nonzero pure-bath part and the first nonzero coupling part of the series (A2). An equivalent statement is that the error must be of higher order than the estimator. To quantify this statement, we take the two-component norm for both $\delta\Omega_{(n)}$ and $\hat{\Omega}_{(n)}$, and demanding

$$\|\delta\Omega_{(n)}\| \equiv \begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \ll \|\hat{\Omega}_{(n)}\| \equiv \begin{pmatrix} \hat{\alpha}_{(n)} \\ \hat{\beta}_{(n)} \end{pmatrix}, \quad (\text{A4})$$

where the comparison is implied for both components and should be understood by comparing the leading powers of the polynomials in α and β . In essence, a “smaller” polynomial should have a larger leading power and thus is of higher order smallness. We refer readers to Appendix ?? for more rigorous descriptions on ordering polynomials by their leading powers. Furthermore, after taking the two-component norm and applying triangular inequality for Eq. (A3), we find,

$$\hat{\beta}_{(n)} - \delta\beta_{(n)} \leq \beta_{(n)} \leq \hat{\beta}_{(n)} + \delta\beta_{(n)}. \quad (\text{A5})$$

According to Eq. (??), the n th-order decoupling condition requires $\beta_{(n)}$ to be an $(n+1)$ th order polynomial in α and β . It now suffices to show that ① $\hat{\beta}_{(n)}$ is an $(n+1)$ th order polynomial and ② the estimator is faithful.

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

$$\hat{\Omega}_{(n)} \equiv (\Omega_{\text{PDD}}^{(1)} + \Omega_{\text{PDD}}^{(2)})(\hat{\Omega}_{(n-1)}) = (\Omega_{\text{PDD}}^{(1)} + \Omega_{\text{PDD}}^{(2)})^n(\Omega). \quad (\text{A6})$$

With $\Omega_{\text{PDD}}^{(1)}, \Omega_{\text{PDD}}^{(2)}$ explicitly defined in Eq. (??) and Eq. (8), and the starting point $\hat{\Omega}_{(0)} \equiv \Omega$, the estimator can shown to be specified by the formula:

$$\begin{aligned} \hat{\Omega}_{(n)} &= \sigma_0 \otimes 4^n B_0 \\ &+ \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B_0}^n(B_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}). \end{aligned} \quad (\text{A7})$$

After taking the two-component norm of the estimator, we have,

$$\begin{pmatrix} \hat{\alpha}_{(n)} \\ \hat{\beta}_{(n)} \end{pmatrix} \simeq \begin{pmatrix} \alpha \\ \alpha^n \beta + \alpha^{n-1} \beta^2 \end{pmatrix}, \quad (\text{A8})$$

where \simeq represents equal the leading powers as described in Appendix ???. This verifies that $\hat{\beta}_{(n)}$ is an $(n+1)$ th order polynomial.

To complete the theory, it now remains to show that $\hat{\Omega}_{(n)}$ is indeed faithful. For simplicity, we make the recognition $\alpha \simeq \beta \simeq \tau_0$. This change does not affect the criteria for n th order decoupling as the polynomial order is unchanged. To show $\delta\alpha_{(n)}$ and $\delta\beta_{(n)}$ are of higher order in τ_0 compared to $\hat{\alpha}_{(n)}$ and $\hat{\beta}_{(n)}$, we claim that:

$$\begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \lesssim \begin{pmatrix} \tau_0^3 \\ \tau_0^{n+2} \end{pmatrix}. \quad (\text{A9})$$

We prove this assertion by mathematical induction. At $n = 1$, i.e. the PDD sequence, the error comes solely from the Magnus series truncation. Taking the two-component norm and applying triangular inequality:

$$\|\delta\Omega_{(1)}\| \leq \sum_{m=3}^{\infty} \|\Omega_{(1)}^{(m)}\| \simeq \|\Omega_{(1)}^{(3)}\| \lesssim \begin{pmatrix} \tau_0^3 \\ \tau_0^3 \end{pmatrix}. \quad (\text{A10})$$

By definition, the higher level error term $\delta\Omega_{(n+1)}$ can be expressed with the lower level functions as:

$$\begin{aligned} \delta\Omega_{(n+1)} &= \Omega_{\text{PDD}}(\Omega_{(n)}) - (\Omega_{\text{PDD}}^{(1)} + \Omega_{\text{PDD}}^{(2)})(\hat{\Omega}_{(n)}) \\ &= \Omega_{\text{PDD}}^{(1)}(\delta\Omega_{(n)}) + \sum_{m=2}^{\infty} \Omega_{\text{PDD}}^{(m)}(\Omega_{(n)}) - \Omega_{\text{PDD}}^{(2)}(\hat{\Omega}_{(n)}), \end{aligned} \quad (\text{A11})$$

where we have used the fact that $\Omega_{\text{PDD}}^{(1)}$ is a linear map. Assuming Eq. (A9) holds for level n , we need to show $\|\delta\Omega_{(n+1)}\| \lesssim (\tau_0^3, \tau_0^{n+3})$. The argument in the original CDD paper by Khodjasteh and Lidar for truncating the Magnus series have only showed that the higher order Magnus term is smaller by proving

$$\|\sum_{m=3}^{\infty} \Omega_{\text{PDD}}^{(m)}(\hat{\Omega}_{(n)})\| \ll \|\hat{\Omega}_{(n+1)}\|.$$

If we take $\delta\Omega_{(n)} = 0$, Eq. (A11) would identity the l.h.s. of above as $\|\delta\Omega_{n+1}\|$, then this inequality would indeed indicate that the error term is of higher order. However Eq. (A11) 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 accumulates from the lower level error $\delta\Omega_{(n)}$. To properly bound the size of the error term, we take the two-component norm on both sides of Eq. (A11):

$$\begin{aligned} \|\delta\Omega_{(n+1)}\| &\lesssim \|\Omega_{\text{PDD}}^{(1)}(\delta\Omega_{(n)})\| + \|\Omega_{\text{PDD}}^{(3)}(\Omega_{(n)})\| \\ &+ \|\Omega_{\text{PDD}}^{(2)}(\hat{\Omega}_{(n)} + \delta\Omega_{(n)}) - \Omega_{\text{PDD}}^{(2)}(\hat{\Omega}_{(n)})\|, \end{aligned} \quad (\text{A12})$$

where the infinite sum of Magnus terms higher than third order is lead by the non-vanishing third order. We now examine the size of these terms. The first term in (A12) is easy,

$$\|\Omega_{\text{PDD}}^{(1)}(\delta\Omega_{(n)})\| = \begin{pmatrix} 4\delta\alpha_{(n)} \\ 0 \end{pmatrix} \simeq \begin{pmatrix} \tau_0^3 \\ 0 \end{pmatrix}.$$

To bound the size of the second term, we make two observations for the PDD Magnus series greater than the third order:

1. At least one $B_{i \neq 0}$ is involved in the coupling part.
2. At least two $B_{i \neq 0}$ is involved in pure bath part.

Therefore, the second term can be bounded by

$$\|\Omega_{\text{PDD}}^{(3)}(\Omega_{(n)})\| \lesssim \begin{pmatrix} \beta_{(n)}^2(\alpha_{(n)} + \beta_{(n)}) \\ \beta_{(n)}(\alpha_{(n)} + \beta_{(n)})^2 \end{pmatrix} \simeq \begin{pmatrix} \tau_0^{2n+2} \\ \tau_0^{n+3} \end{pmatrix},$$

where we have used $\alpha_{(n)} \lesssim \hat{\alpha}_{(n)}$ and $\beta_{(n)} \lesssim \hat{\beta}_{(n)}$ according to the induction hypothesis. This can be alternatively shown by explicitly calculating the third order Magnus term using Eq. (??). To bound the third term in (A12), we can calculate the difference $\Omega_{\text{PDD}}^{(2)}(\hat{\Omega} + \delta\Omega) - \Omega_{\text{PDD}}^{(2)}(\hat{\Omega})$ using the expression for $\Omega_{\text{PDD}}^{(2)}$ from Eq. (8). In the end,

$$\begin{aligned} & \left| \left| \Omega_{\text{PDD}}^{(2)}(\hat{\Omega} + \delta\Omega) - \Omega_{\text{PDD}}^{(2)}(\hat{\Omega}) \right| \right| \lesssim \begin{pmatrix} 0 \\ \delta\alpha_{(n)}\delta\beta_{(n)} \end{pmatrix} \\ & + \begin{pmatrix} 0 \\ \hat{\beta}_{(n)}\delta\alpha_{(n)} + \hat{\alpha}_{(n)}\delta\beta_{(n)} + \hat{\beta}_{(n)}\delta\beta_{(n)} \end{pmatrix} \simeq \begin{pmatrix} 0 \\ \tau_0^{n+3} \end{pmatrix}. \end{aligned}$$

Since all three terms are bounded by $(\tau_0^3, \tau_0^{n+3})^\top$, so is

their sum. This proves our assertion Eq. (A9). \square

Having showed its faithfulness, we are now confident in using the estimator Eq. (A2) to study the full $\Omega_{(n)}$ of CDD_n. In particular, we have $\alpha_{(n)} \approx \hat{\alpha}_{(n)} = 4^n\alpha$ and $\beta_{(n)} \approx \hat{\beta}_{(n)}$. Similar as the PDD case, we can derive a strict bound and an approximate bound. For the strict bound, we have

$$\begin{aligned} \hat{\beta}_{(n)} & \leq 2^{n(n+1)} \|\text{ad}_{B_0}^n(B_1)\| \\ & \quad + 2^{n^2} \|\text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\})\| \\ & \leq 2^{n(n+1)} \alpha^{n-1} \beta (2^n \alpha + \alpha + \beta). \end{aligned} \quad (\text{A13})$$

-
- [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).