

# Efficacy of noisy dynamical decoupling

Jiaan Qi,<sup>1</sup> Xiansong Xu,<sup>2</sup> Dario Poletti,<sup>2</sup> and Hui Khoon Ng<sup>1,3,4</sup>

<sup>1</sup>*Yale-NUS College, Singapore*

<sup>2</sup>*Singapore University of Technology and Design*

<sup>3</sup>*Centre for Quantum Technologies, National University of Singapore*

<sup>4</sup>*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. 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. We illustrate a typical sequence of DD control gates in Figure 1.

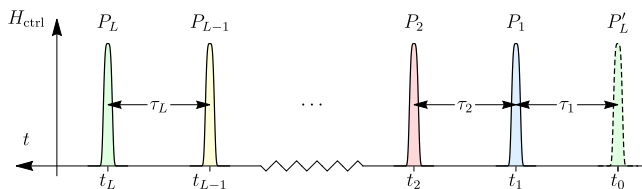


FIG. 1: Illustration of the external control pulses for a typical DD sequence. Different gates are filled with different colors. The dashed gate at  $t_0$  corresponds to the previous DD period.

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 completion of the sequence. Different strategies could differ in the following aspects: (i) the transformation group  $\mathbf{G}$ .—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  $\tau$  in between consecutive gates. A nonzero  $\tau$  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 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.  $\Omega_{DD}$  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  $\tau H$ , 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  $\|\Omega_{\text{DD}}^{(n)}\| \sim \|\tau H\|^n$ , compared to  $\|\Omega_{\text{DD}}^{(1)}\| \sim \|\tau H\|$  without DD.

## B. 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. 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  $\Omega$  and  $\rho_{\text{B}}$ , the transformation on  $\Omega$  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:

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

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

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

where we use the notation  $(\alpha, \beta)^{n+1}$  to represent a polynomial that is  $(n+1)$ th order in  $\alpha$  and  $\beta$ . Provided that the Magnus series converges absolutely,  $\beta_{\text{DD}}$  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  $U(0, T) \equiv e^{-iT H_{\text{eff}}}$ , for some effective time-independent system-bath Hamilton operator  $H_{\text{eff}}$ .  $H_{\text{eff}}$  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,

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

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  $H_{\text{eff},\text{B}}$  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

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

(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  $\mathcal{H}_{\text{S}} \otimes \mathcal{H}_{\text{B}}$ :

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

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

### 3. Channel fidelity measures

## C. PDD and its concatenation

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  $\tau$  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  $\Omega_{DD}$  of Eqs. (2) and (3), now re-labeled as  $\Omega_{PDD}$ , can be worked out to be

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

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

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

where  $[\cdot, \cdot]$  and  $\{\cdot, \cdot\}$  are the commutator and anti-commutator, respectively.  $\Omega_{PDD}$  should be compared against the evolution operator (over the same time period of  $4\tau$ ) 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  $\Omega_{PDD}^{(1)}$ —usually the dominant term—with  $\Omega$ , we see that the  $H_{SB}$  in  $\Omega$  no longer appears in  $\Omega_{PDD}^{(1)}$  and  $\Omega_{PDD}^{(1)}$  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  $\Omega_{PDD}^{(2)}$ , 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  $\Omega$ -like quantities into  $\tau H$ s (except when we talk about the Magnus series) – it is important to keep the time parameter visible. Sometimes,  $\tau$  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  $\Omega$ . This becomes especially important when considering CDD.]

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)  $\tau_0$ , 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  $\tau_0$  or  $\tau_n$ . CDD at level  $n$ , denoted as  $CDD_n$ , is then a sequence of  $L_n \equiv L^n$  pulses separated by time interval  $\tau_0$  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.

### 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,

$$\Phi_{\text{SB}} \leq \phi_{\text{SB}}. \quad (20)$$

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:

$$\Phi_{\text{SB}} \simeq \|\Omega_{\text{PDD}}^{(2)}\| \leq \phi_{\text{SB}}. \quad (21)$$

Using the expression for  $\Omega_{\text{PDD}}^{(2)}$  from Eq. (14), we have

$$\begin{aligned} \|\Omega_{\text{PDD}}^{(2)}\| &\leq 2\tau^2(2\|[B_I, B_X]\| + \|[B_I, B_Y]\| + \|[B_X, B_Z]\|) \\ &\leq 4\phi_B\tau(2\|B_X\| + \|B_Y\|) + 4\tau^2\|B_X\|\|B_Z\| \\ &\leq 12\phi_B\phi_{\text{SB}} + 4\phi_{\text{SB}}^2 \end{aligned} \quad (22)$$

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 (21), it then suffices to require

$$12\phi_B + 4\phi_{\text{SB}} \leq 1. \quad (23)$$

This amounts to a requirement that both  $\phi_{\text{SB}}$  and  $\phi_B$  be small for PDD to work well. That  $\phi_{\text{SB}}$  has to be small comes as no surprise— $\phi_{\text{SB}}$  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  $\phi_B$  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  $\phi_B$  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.  $\phi_B$  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  $\phi_B$ . Tracing back to the original calculation, we see that part of the pre-factor of 12 of  $\phi_B$  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  $\phi_B$  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 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)

$$\left\|\sum_{i=1}^3 \sigma_i \otimes B_i\right\| \approx \sqrt{\sum_{i=1}^3 \|B_i\|^2}. \quad (24)$$

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

$$\begin{aligned} \beta_{\text{PDD}}^2 &\approx \|-4i[B_0, B_1]\|^2 + \|-2i[B_0, B_2] - 2\{B_1, B_3\}\|^2 \\ &\leq (8\alpha\|B_1\|)^2 + (4\alpha\|B_2\| + 4\|B_1\|\|B_3\|)^2, \end{aligned} \quad (25)$$

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

$$\max_{x_1^2+x_2^2+x_3^2=1} (2\alpha x_1)^2 + (\alpha x_2 + \beta x_1 x_3)^2 \leq \frac{1}{16}. \quad (26)$$

Solution of this constrained optimization problem leads



to:

$$\left(\frac{\beta}{\sqrt{3}} \leq \alpha \leq \frac{1}{8}\right) \text{ or } \left(10\alpha^2 + \frac{9\alpha^4}{\beta^2} + \beta^2 \leq \frac{1}{4}\right). \quad (27)$$

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 (27) each determines a region in the  $(\alpha, \beta)$  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:

$$\alpha + \beta \leq \frac{1}{4}. \quad (28)$$

To examine the accuracy of the two theoretical bounds, we perform numeric tests as a verification. We first randomly generate four  $2 \times 2$  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  $\Omega_{\text{PDD}}$  and error phase  $\beta_{\text{PDD}}$ . For each fixed  $(\alpha, \beta)$  point, we numerically maximize the error phase reduction ratio  $\beta_{\text{PDD}}/\beta$  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 (27), together with the absolute convergence region (28) 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” (27), 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 (28), 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.

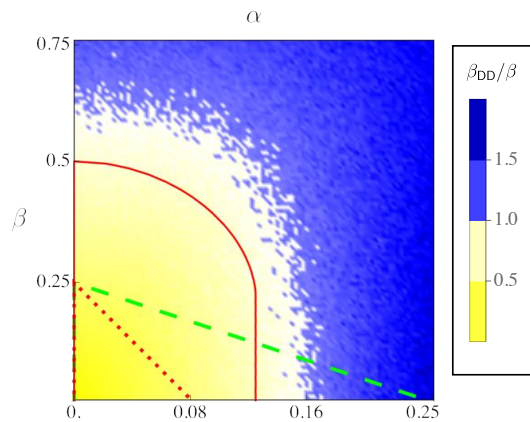


FIG. 2: The noise reduction ratio  $\beta_{\text{DD}}/\beta$  obtained through maximization over random bath operators plotted in the  $(\alpha, \beta)$  parameter plane. In comparison, the theoretical boundaries for noise reduction are also plotted, with dotted red line for (??) and solid red line for (27), also with the boundary for the absolute-convergence condition (28) in dashed green line. Note the ranges for  $\alpha$  and  $\beta$  are chosen differently for better visual effect.

#### IV. NOTE ON 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:

$$H(t) = H_e + H_c(t), \quad t_{i-1} < t \leq t_i, \quad (29)$$

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. (29) 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:

$$\tilde{U}(t_i, t_{i-1}) = P_i e^{-i\tilde{\Omega}_i}, \quad (30)$$

where the gate-dependent  $\tilde{\Omega}_i$  replaces the free evolution generator  $\Omega$  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,

$$\tilde{U}_{\text{DD}} = e^{-iG_L \tilde{\Omega}_L G_L^\dagger} \dots e^{-iG_2 \tilde{\Omega}_2 G_2^\dagger} e^{-iG_1 \tilde{\Omega}_1 G_1} \equiv e^{-i\tilde{\Omega}_{\text{DD}}}. \quad (31)$$

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  $\eta$ , which is defined later, to quantify the noise strength associated with the gates. The break-even condition is, in essence, to invert the inequality

$$\tilde{\Phi}_{\text{SB}}(\phi_B, \phi_{\text{SB}}, \eta) < \phi_{\text{SB}}, \quad (32)$$

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  $\eta$ , 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:

$$\tilde{U}(t_i, t_{i-1}) = \tilde{P}_i e^{-i\Omega}, \quad (33)$$

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:

$$\tilde{P}_i = \exp[-i(\Omega_{P_i} + \eta \Gamma_{P_i})], \quad (34)$$

where  $\Omega_{P_i}$  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  $\eta$  to facilitate order tracking and normalized  $\Gamma_{P_i}$ . For PDD in particular, the sequence comprises only  $X$  and  $Z$  gates. And we may write the noisy  $X$  and  $Z$  pulses as:

$$\tilde{X} = e^{-i(\frac{\pi}{2}\sigma_1 + \eta \Gamma')}, \quad \tilde{Z} = e^{-i(\frac{\pi}{2}\sigma_3 + \eta \Gamma'')}, \quad (35)$$

where  $\Gamma'$  and  $\Gamma''$  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

$$\begin{aligned} \tilde{U}_{\text{DD}} &= \tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \\ &= e^{-i\Gamma_3} e^{-i\Omega_3} e^{-i\Gamma_2} e^{-i\Omega_2} e^{-i\Gamma_1} e^{-i\Omega_1} e^{-i\Gamma_0} e^{-i\Omega_0}, \end{aligned} \quad (36)$$

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  $\eta$ , we can derive

$$\begin{aligned} \Gamma_3 &= \eta(\sigma_0 \otimes B''_0 - \frac{2}{\pi} \sigma_1 \otimes B'_2 + \frac{2}{\pi} \sigma_2 \otimes B'_1 + \sigma_3 \otimes B'_3), \\ \Gamma_2 &= \eta(\sigma_0 \otimes B'_0 - \sigma_1 \otimes B'_1 + \frac{2}{\pi} \sigma_2 \otimes B'_3 + \frac{2}{\pi} \sigma_3 \otimes B'_2), \\ \Gamma_1 &= \eta(\sigma_0 \otimes B''_0 + \frac{2}{\pi} \sigma_1 \otimes B'_2 + \frac{2}{\pi} \sigma_2 \otimes B'_1 - \sigma_3 \otimes B'_3), \\ \Gamma_0 &= \eta(\sigma_0 \otimes B'_0 + \sigma_1 \otimes B'_1 + \frac{2}{\pi} \sigma_2 \otimes B'_3 - \frac{2}{\pi} \sigma_3 \otimes B'_2). \end{aligned} \quad (37)$$

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

Every term on the exponent in Eq. (36) is now at least linear in the smallness  $\eta$  or  $\|\Omega\|$  and we can invoke the Magnus formula to calculate  $\tilde{\Omega}_{\text{PDD}}$ . The first order term suggests:

$$\tilde{\Omega}_{\text{DD}}^{(1)} = \sigma_0 \otimes (4B_0 + 2\eta(B'_0 + B''_0)) + \sigma_2 \otimes \frac{4}{\pi} \eta(B'_1 + B'_3). \quad (38)$$

When the gate noise strength  $\eta$  is at least of similar size as the environmental noise  $\phi_{\text{SB}}$ —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  $\eta$ . we upper-bound the size of the first order coupling by:

$$\tilde{\Phi}_{\text{SB}} \approx \left\| \frac{4\eta}{\pi} (B'_1 + B'_3) \right\| \leq \frac{8}{\pi} \eta. \quad (39)$$

This results in the first order breakeven condition:

$$\eta \leq \frac{\pi}{8} \phi_{\text{SB}}. \quad (40)$$

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  $\eta$  and up to second order in the smallness  $\phi_{\text{B}}$  or  $\phi_{\text{SB}}$  (alternatively: second order in  $\phi$ ). Focusing on the interaction part, the result becomes:

$$\tilde{\Omega}_{\text{DD:SB}} = \Omega_{\text{DD:SB}} + \sigma_2 \otimes \frac{4}{\pi} \eta (B_1'' + B_3') + \mathcal{O}(\eta\phi, \eta^2). \quad (41)$$

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

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

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 3. 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 3a); and the relation between the maximally tolerable gate noise strength versus the environmental noise strength (Figure 3b). 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. (38), 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  $\eta$ . Hence an alternative condition should be considered. The revised first

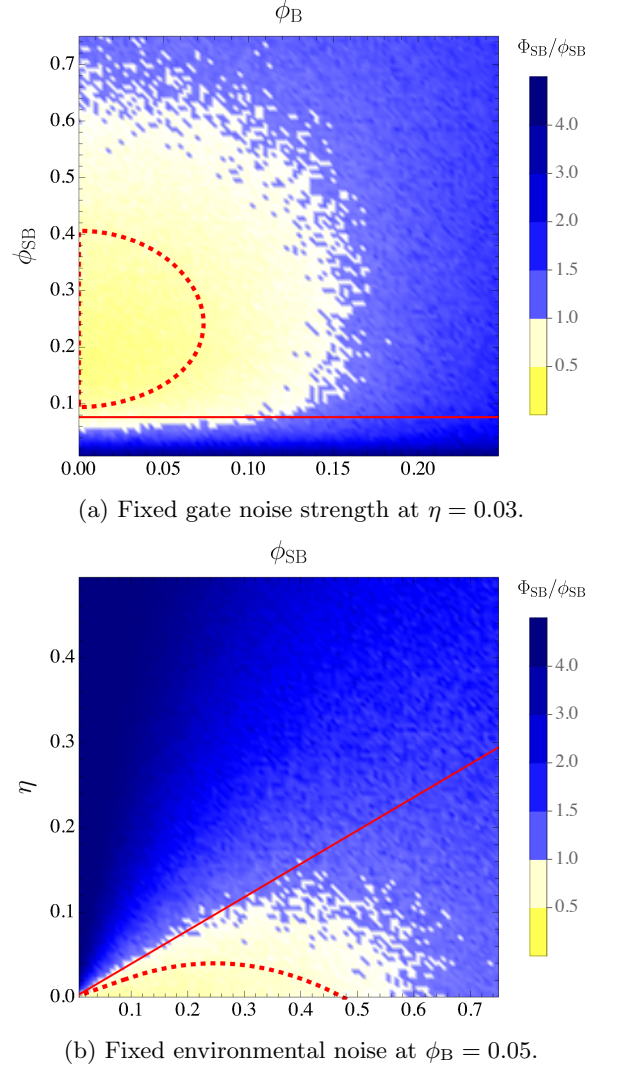


FIG. 3: The noise reduction diagrams for noisy PDD with either fixed gate noise strength or environmental noise strength. The first order and second conditions derived in the main text are indicated with red solid lines and red dashed line respectively for comparison.

order condition demands:

$$\tilde{\Omega}_{\text{DD:SB}}^{(1)} = \mathcal{O}(\eta^2). \quad (43)$$

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  $\Gamma_i$  is at most on the order of the smallness parameter  $\eta$ . Hence,

$$\tilde{P}_i = P_i e^{-i\eta\Gamma + \mathcal{O}(\eta^2)}, \quad \forall i. \quad (44)$$

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

$$e^{-i\tilde{\Omega}} = e^{-i\eta\Gamma} e^{-i\Omega}. \quad (45)$$

Since  $\Gamma$  and  $\Omega$  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  $\eta$ . The already established map  $\Omega \rightarrow \Omega_{\text{DD}}$  for the idea-pulse DD sequence can then be reused to estimate  $\Omega_{\text{DD}}$ . 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  $(\tilde{\phi}_B, \tilde{\phi}_{\text{SB}}) \equiv (\|\tilde{\Omega}_B\|, \|\tilde{\Omega}_{\text{SB}}\|)$ :

$$\begin{cases} \tilde{\phi}_B \lesssim (\eta + \phi_B)(1 + \eta + \phi_{\text{SB}}) \equiv \tilde{\phi}_B^{(\text{ub})} \\ \tilde{\phi}_{\text{SB}} \lesssim (\eta + \phi_{\text{SB}})(1 + \eta + \phi_B) \equiv \tilde{\phi}_{\text{SB}}^{(\text{ub})} \end{cases} \quad (46)$$

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

$$\tilde{\Phi}_{\text{SB}} \leq \Phi_{\text{SB}}(\tilde{\phi}_B^{(\text{ub})}, \tilde{\phi}_{\text{SB}}^{(\text{ub})}) < \phi_{\text{SB}}. \quad (47)$$

Further bounding the right hand side by  $\phi_{\text{SB}}$  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

$$\tilde{P}_i = P_i e^{-i\theta \cdot \sigma}, \quad \forall i. \quad (48)$$

We use the notation  $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ , and  $\theta \equiv (\theta_1, \theta_2, \theta_3)$  with  $\theta_i$  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:

$$\begin{cases} \tilde{\phi}_B \leq \phi_B + \frac{1}{3}\theta\phi_{\text{SB}}^2, \\ \tilde{\phi}_{\text{SB}} \leq \phi_{\text{SB}} + \theta^2\phi_{\text{SB}} \end{cases}, \quad (49)$$

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:

$$2\phi_B^2 + \frac{(\theta + \phi_{\text{SB}})^4}{4\phi_{\text{SB}}^2} \leq \frac{1}{16}, \quad (50)$$

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

$$\theta \leq (1 - 32\phi_B^2)^{\frac{1}{4}} \sqrt{\frac{\phi_{\text{SB}}}{2}} - \phi_{\text{SB}} \lesssim \sqrt{\frac{\phi_{\text{SB}}}{2}}. \quad (51)$$

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  $\phi_{\text{SB}}$  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  $\tau_p$ , 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:

$$\tilde{U}(t_i, t_{i-1}) = e^{-i(\Omega_{P_i} + \epsilon\Gamma_i + (\tau_p/\tau)\Omega)} e^{-i(1 - (\tau_p/\tau))\Omega}. \quad (52)$$

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. (34) 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. (38) to derive the first order term, we can estimate the error phase and consequently derivative the breakeven condition as

$$\tilde{\Phi}_{\text{SB}} \lesssim \frac{8}{\pi} \left( \frac{\tau_p}{\tau} \right) \phi_{\text{SB}} + \frac{8}{\pi} \epsilon \leq \phi_{\text{SB}}. \quad (53)$$

The ratio  $\tau_p/\tau$  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

$$\tau_p/\tau \leq \frac{\pi}{8} \approx 0.39, \quad (54)$$

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:

$$H(t) = H_g(t) + H_e + H_{ge}(t), \quad 0 \leq t \leq \tau, \quad (55)$$

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

$$U(\tau) = P_g e^{-i\tilde{\Omega}_g}, \quad (56)$$

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:

$$\tilde{\Omega}_g \approx \int_0^\tau U_g^\dagger(t) H_e U_g(t) dt + \int_0^\tau U_g^\dagger(t) H_{ge}(t) U_g(t) dt. \quad (57)$$

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:

$$\begin{aligned} \int_0^\tau U_X^\dagger(t) H_e U_X(t) dt &\equiv \sigma_0 B_0 + \sigma_1 B'_1 + \sigma_2 B'_2 + \sigma_3 B'_3, \\ \int_0^\tau U_Z^\dagger(t) H_e U_Z(t) dt &\equiv \sigma_0 B_0 + \sigma_1 B''_1 + \sigma_2 B''_2 + \sigma_3 B''_3, \end{aligned} \quad (58)$$

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

$$\left\| \int_0^\tau U_g^\dagger(t) H_{ge}(t) U_g(t) dt \right\| \leq \int_0^\tau \|H_{ge}(t)\| dt \equiv \eta_{ge}, \quad (59)$$

where the small number  $\eta_{ge}$  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  $\Gamma_g$  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:

$$\begin{aligned} \tilde{\Omega}_{\text{DD}}^{(1)} &= \sigma_0 \otimes (4B_0 + 2\eta_{ge}(A'_0 + A''_0)) \\ &\quad + \sigma_2 \otimes (2(B'_2 - B''_2) + 2\eta_{ge}(A'_2 - A''_2)). \end{aligned} \quad (60)$$

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

$$\tilde{\Phi}_{\text{SB}} \lesssim 2\|B'_2 - B''_2\| + 4\eta_{ge}. \quad (61)$$

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

$$H_g(t) = \frac{1}{2}\omega_g(t)\sigma_g, \quad g \in \{X, Z\}, \quad (62)$$

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  $\tau$  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:

$$\begin{aligned} \int_0^\tau U_g^\dagger(t) H_e U_g(t) dt &= \sigma_0 B_0 + \\ &\begin{cases} \sigma_1 B_1 + c_X(\sigma_2 B_2 + \sigma_3 B_3) + s_X(\sigma_2 B_3 - \sigma_3 B_2), & g = X; \\ \sigma_3 B_3 + c_Z(\sigma_1 B_1 + \sigma_2 B_2) + s_Z(\sigma_1 B_2 - \sigma_2 B_1), & g = Z, \end{cases} \end{aligned} \quad (63)$$

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

$$c_g \equiv \frac{1}{\tau} \int_0^\tau \cos(\theta_g(t)) dt, \quad s_g \equiv \frac{1}{\tau} \int_0^\tau \sin(\theta_g(t)) dt. \quad (64)$$

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. (58) is still zero, which is all that needed for first order decoupling. Focusing on the difference in the  $B_2$  component, we have

$$\begin{aligned} \|B'_2 - B''_2\| &= \|(c_X - c_Z)B_2 + s_X B_3 + s_Z B_1\| \\ &\leq (|c_X - c_Z| + |s_X| + |s_Z|) \cdot \phi_{\text{SB}}. \end{aligned} \quad (65)$$

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

$$\omega_g(t) = \begin{cases} 0, & 0 \leq t < \tau' \\ \pi/\tau_p, & \tau' \leq t < \tau' + \tau_p \\ 0, & \tau' + \tau_p \leq t \leq \tau \end{cases} \quad (66)$$

It is straightforward to calculate that  $s_g = \frac{2\tau_p}{\pi\tau}$ . Hence the resulting error phase is independent of  $\tau'$ . After substituting into Eq. (61), we arrive at the breakeven condition

$$\tilde{\Phi}_{\text{SB}} \lesssim \frac{8}{\pi} \left( \frac{\tau_p}{\tau} \right) \phi_{\text{SB}} + 4\eta_{ge} \leq \phi_{\text{SB}}. \quad (67)$$

This produces the same behavior compared to Eq. (53), apart from numerical factors arisen from difference in gate error definitions.

## V. SCALING UP PROTECTION – THE ACCURACY THRESHOLD

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

$$\text{Memory setting: } \epsilon_{\text{DD}} < \epsilon(L), \quad (68)$$

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

$$\text{Computational setting: } \epsilon_{\text{DD}} < \epsilon(1), \quad (69)$$

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

These two perspectives result in very different overall decoupling strategies. Especially if one is using CDD to “scale up protection”. As illustrated in Fig. 4, for the quantum memory setting, as total time is fixed  $T$ , one only needs to come up with finer and finer slicing of the time evolution pieces; but for the quantum computation setting, one needs 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 serious 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 (69) will yield a requirement on

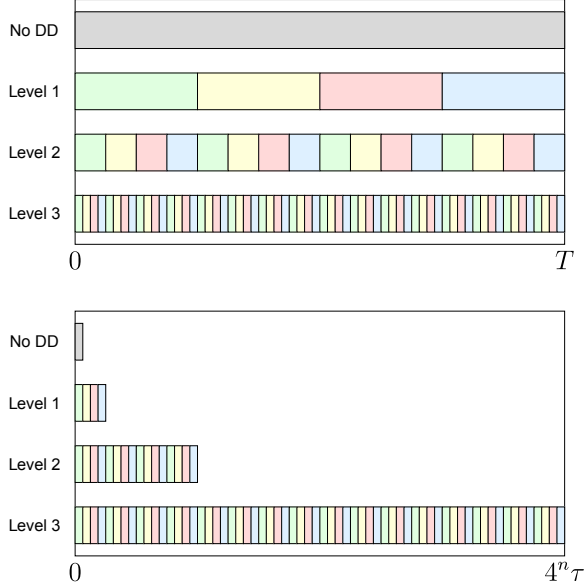


FIG. 4: An illustration of the CDD time evolution maps under increasing concatenation levels. We compare the quantum memory setting (above) to the quantum computation setting (under).

the noise parameters characterizing the noise in the system and the DD pulses. We refer to that requirement as the “break-even point” for DD, borrowing terminology from quantum error correction and fault-tolerant quantum computing. (There was a comment in the .tex file that some past papers have analyzed the memory case. What are those papers? Do you just mean the Khodjasteh-Lidar 2007 paper? That paper looks only at imperfections due to finite-width pulses, not control errors as we do here. Their paper hence only has a single-parameter  $\Phi_{\text{SB}}$ , not our two-parameter  $\Phi_{\text{SB}}$  and  $\eta$  situation. In fact, in current experiments  $\eta$  is likely the larger problem; memory noise is usually much smaller than gate noise. Or are there other earlier papers?) (Yes, that paper in particular. Their conclusion that CDD decouples to order  $n$  and dramatically outperforms PDD come from this perspective. Our paper is focused on the quantum computation case which is more practical. See my new added picture 4). And let me quote their words (bottom left page 10) on how they do the comparison: “The higher order data for PDD are obtained by simply repeating the basic sequence and shrinking the pulse interval. This is necessary to get the long-term behavior of decoupling. ...” I move this memory/computation argument from the previous section as it is more appropriate when we discuss the scaling behavior. —Jiaan)

(To be merged with the new text.) The conditions for achieving noise suppression are already derived previously for the idealistic PDD and CDD case. In general, we find that smaller  $(\alpha, \beta)$  values would lead to to bet-

ter 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.

Next, we turn to the situation of CDD. We can certainly discuss the break-even point for a particular  $\text{CDD}_n$  sequence, treating it simply as a “flattened” sequence of pulses, and ask when the error phase after  $\text{CDD}_n$  is smaller than  $\phi_{\text{SB}}$  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 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 parameter 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  $\eta < \eta_0$ , such that

$$\Phi_{\text{SB},n+1} < \Phi_{\text{SB},n} \quad \forall n = 1, 2, \dots \quad (70)$$

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

If a noise threshold indeed exist, then ideal case, where the noise is non-existent, would definitely apply. We first consider CDD with perfect pulses. Khodjasteh and Lidar propose in their paper to truncate to the second order term gave in their paper an estimation of the  $\text{CDD}_n$  generator as:

$$\begin{aligned}\Omega_{(n)} &\approx \sigma_0 \otimes 4^n B_0 \\ &+ \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B_0}^n(B_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}).\end{aligned}\quad (71)$$

It is shown in the appendix B that the above expression indeed captures the leading order behavior of  $\Omega_{(n)}$ . Different from the leading order term in the Magnus series, the leading order “behavior” is manifested separately for the pure bath part and the interaction part. As mentioned in the Preliminaries section,  $\text{CDD}_n$  achieves  $n$ th-order decoupling. From the leading order expression, we see it is indeed so as the interaction part is of  $(n+1)$ th order. From Eq. (71), we can estimate  $\Phi_{B,n} \simeq 4^n \phi_B$  and derive an upper bound for the error phase:

$$\Phi_{\text{SB},n} \lesssim 2^{n(n+2)} \phi_B^n \phi_{\text{SB}} f_n\left(\frac{\phi_{\text{SB}}}{\phi_B}\right), \quad (72)$$

where we define the piece-wise function:

$$f_n(x) = \begin{cases} 1, & x \leq \sqrt{4^n - 1}, \\ \frac{1}{2^n} \sqrt{\left(\frac{4^n - 1}{2x} + \frac{x}{2}\right)^2 + 1}, & x > \sqrt{4^n - 1}. \end{cases} \quad (73)$$

to remind the reader,  $\phi_B \equiv \phi_B(\tau_0) \equiv \tau_0 \|H_B\|$ , and  $\phi_{\text{SB}} \equiv \phi_{\text{SB}}(\tau_0) \equiv \tau_0 \|H_{\text{SB}}\|$ , noting the dependence of both quantities on  $\tau_0$ , 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. (B5)]. For use in Eq. (70), 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,  $\text{CDD}_n$  achieves  $n$ th-order decoupling. This can be understood from the error phase expression:  $\Phi_{\text{SB},n}$  after  $\text{CDD}_n$  is of order  $\phi^{n+1}$ , for  $\phi \sim \phi_B, \phi_{\text{SB}}$ . That  $\phi_B$  enters the error phase should again be of no surprise—as in the PDD case,  $\phi_B$  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  $\Phi_{\text{SB},n}$ , 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  $\tau$  is the experimental limit for the switch time between consecutive

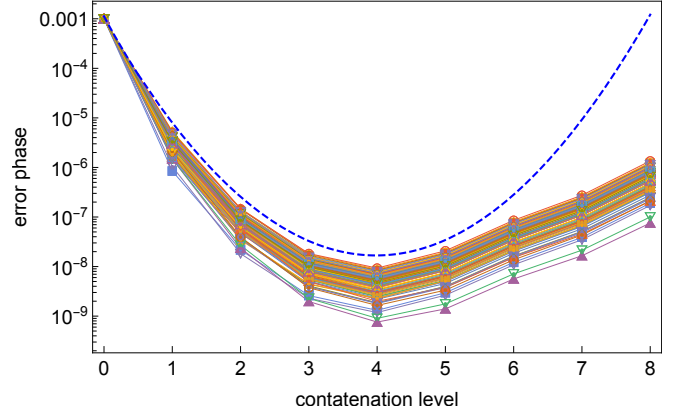


FIG. 5: The error phase  $\phi_{\text{SB},n}$  as a function of concatenation level  $n$ . The thin color lines are obtained from numerically simulating  $\text{CDD}_n$  with randomly generated bare Hamiltonian operators satisfying  $\phi_B = \phi_{\text{SB}} = 0.001$ . The blue dashed line is our theoretical upper bound. For this configuration, we set the maximal concatenation level at 4.

pulses), such factors mean that the error phase eventually increases for large enough  $n$ : The exponentially decreasing  $\phi^n$  factor is eventually overcome by the super-exponentially increasing  $2^{n^2}$  factor. This means there is no accuracy threshold, i.e., there is no level of noise that is weak enough such that  $\Phi_{\text{SB},n+1} \leq \Phi_{\text{SB},n}$  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. 5 plots this situation of fixed  $\tau_0$ , 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 of it, we examine the recursive relation between concatenation levels, which suggests

$$\begin{aligned}B_{n+1,0} &= 4B_{n,0} \\ B_{n+1,1} &= -4i[B_{n,0}, B_{n,1}] \\ B_{n+1,2} &= -2i[B_{n,0}, B_{n,1}] \\ B_{n+1,3} &= 0\end{aligned}\quad (74)$$

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:

$$n_{\text{max}} \leq \lceil -\log_4(\phi_B) - 3/2 \rceil, \quad (75)$$

where  $\lceil \cdot \rceil$  is the ceiling operator.

Note that the existence of a maximal useful concatenation level does not technically contradict the statement that higher-order decoupling is achieved by higher level

CDD. Using decoupling order to quantify the degree of noise removal requires, in the first place, the convergence of the Magnus series, so that the  $(n+1)$ th- and higher-order terms in the series are a small correction to the first  $n$  terms. However, if the DD scheme is designed such that the total time for the sequence grows exponentially with  $n$ , as is the case for CDD if  $\tau_0$  is fixed as  $n$  increases, eventually, we exceed the convergence criterion and the decoupling order stops being a reasonable indicator of successful noise removal.

We can, however, analyze a different situation: to have fixed  $\tau_n \equiv T$ , so that the CDD sequence takes the same amount of time, regardless of  $n$ . This requires  $\tau_0 = \frac{T}{4^n}$  for each  $n$ , so that the pulses are applied at shorter and shorter time intervals as the concatenation level increases. This can be the practical choice if one is doing computation where computational gates, which can only be applied at the end of a complete DD sequence in order to not interfere with the noise averaging process, have to be applied at a particular clock rate. Increasing  $n$  to increase the noise removal capabilities must not increase the total time taken for the DD sequence. 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  $\phi_B$  and  $\phi_{SB}$ , in our estimate of the CDD error phases, so that that the  $n$  dependence in  $\tau_0$  is explicit. (I suggest we analyze this case as well, to see if an accuracy threshold exists. Jiaan - 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  $\tau_0$  can be in an experiment.

## B. Noisy pulses

(To add.)  
(To add.)

## 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 (37)

Here we explain the detailed steps to derive Eqs. (37). 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  $\Gamma_i$  to the leading order in the smallness  $\eta$ . 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:

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

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  $\Gamma_i$ :

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

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

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

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

The expressions in Eqs. (A4) can be further calculated by writing out the series expansion for the leading order terms in  $\eta$ . Let us focus on  $\Gamma_3$ , 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  $\Gamma''$ , 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  $\Gamma_3$  in Eq. (37). One can proceed to verify  $\Gamma_2$ ,  $\Gamma_1$  and  $\Gamma_0$  using similar steps.

## Appendix B: Analyzing CDD<sub>n</sub>

(To be edited.) It was suggested by the authors of CDD that the decoupling order for CDD<sub>n</sub> 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<sub>n</sub>, it is infeasible to flatten out the gates and calculate the combined Magnus series directly. Instead it is easier to study the iteration map

$$\Omega_{(n)} = \Omega_{\text{PDD}}(\Omega_{(n-1)}) = \sum_{m=1}^{\infty} \Omega_{\text{PDD}}^{(m)}(\Omega_{(n-1)}), \quad (\text{B1})$$

where the update map  $\Omega_{\text{PDD}}$  is defined in Eq. (??). Backtracking the map from level  $n$  to level 0, we have

$\Omega_{(n)} = \Omega_{\text{PDD}} \circ \cdots \circ \Omega_{\text{PDD}}(\Omega) \equiv (\Omega_{\text{PDD}})^n(\Omega)$ . In the end, the effective Hamiltonian  $\Omega_{(n)}$  can be expanded as:

$$\Omega_{(n)} = \sum_{m=1}^{\infty} \Omega_{(n)}^{(m)}(\Omega), \quad (\text{B2})$$

where  $\Omega_{(n)}^{(m)}$ , with the size of  $\mathcal{O}(\|\Omega\|^m)$ , is the  $m$ th order Magnus term in  $\Omega$ . Notice that this series is different from the series in Eq. (B1), 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  $\Omega$ , 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

$$\Omega_{(n)} = \hat{\Omega}_{(n)} + \delta\Omega_{(n)}. \quad (\text{B3})$$

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 (B2). 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

$$\|\delta\Omega_{(n)}\| \equiv \begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \ll \|\hat{\Omega}_{(n)}\| \equiv \begin{pmatrix} \hat{\alpha}_{(n)} \\ \hat{\beta}_{(n)} \end{pmatrix}, \quad (\text{B4})$$

where the comparison is implied for both components and should be understood by comparing the leading powers of the polynomials in  $\alpha$  and  $\beta$ . 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. (B3), we find,

$$\hat{\beta}_{(n)} - \delta\beta_{(n)} \leq \beta_{(n)} \leq \hat{\beta}_{(n)} + \delta\beta_{(n)}. \quad (\text{B5})$$

According to Eq. (6), the  $n$ th-order decoupling condition requires  $\beta_{(n)}$  to be an  $(n+1)$ th order polynomial in  $\alpha$  and  $\beta$ . 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:

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

With  $\Omega_{\text{PDD}}^{(1)}, \Omega_{\text{PDD}}^{(2)}$  explicitly defined in Eq. (??) and Eq. (14), and the starting point  $\hat{\Omega}_{(0)} \equiv \Omega$ , the estima-

tor can shown to be specified by the formula:

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

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

$$\begin{pmatrix} \widehat{\alpha}_{(n)} \\ \widehat{\beta}_{(n)} \end{pmatrix} \simeq \begin{pmatrix} \alpha \\ \alpha^n \beta + \alpha^{n-1} \beta^2 \end{pmatrix}, \quad (\text{B8})$$

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

To complete the theory, it now remains to show that  $\widehat{\Omega}_{(n)}$  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  $\tau_0$  compared to  $\widehat{\alpha}_{(n)}$  and  $\widehat{\beta}_{(n)}$ , we claim that:

$$\begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \lesssim \begin{pmatrix} \tau_0^3 \\ \tau_0^{n+2} \end{pmatrix}. \quad (\text{B9})$$

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:

$$\|\delta\Omega_{(1)}\| \leq \sum_{m=3}^{\infty} \|\Omega_{(1)}^{(m)}\| \simeq \|\Omega_{(1)}^{(3)}\| \lesssim \begin{pmatrix} \tau_0^3 \\ \tau_0^3 \end{pmatrix}. \quad (\text{B10})$$

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

$$\begin{aligned}\delta\Omega_{(n+1)} &= \Omega_{\text{PDD}}(\Omega_{(n)}) - (\Omega_{\text{PDD}}^{(1)} + \Omega_{\text{PDD}}^{(2)})(\widehat{\Omega}_{(n)}) \\ &= \Omega_{\text{PDD}}^{(1)}(\delta\Omega_{(n)}) + \sum_{m=2}^{\infty} \Omega_{\text{PDD}}^{(m)}(\Omega_{(n)}) - \Omega_{\text{PDD}}^{(2)}(\widehat{\Omega}_{(n)}),\end{aligned}\quad (\text{B11})$$

where we have used the fact that  $\Omega_{\text{PDD}}^{(1)}$  is a linear map. Assuming Eq. (B9) 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

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

If we take  $\delta\Omega_{(n)} = 0$ , Eq. (B11) 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. (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 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. (B11):

$$\begin{aligned}\|\delta\Omega_{(n+1)}\| &\lesssim \|\Omega_{\text{PDD}}^{(1)}(\delta\Omega_{(n)})\| + \|\Omega_{\text{PDD}}^{(3)}(\Omega_{(n)})\| \\ &+ \|\Omega_{\text{PDD}}^{(2)}(\widehat{\Omega}_{(n)} + \delta\Omega_{(n)}) - \Omega_{\text{PDD}}^{(2)}(\widehat{\Omega}_{(n)})\|,\end{aligned}\quad (\text{B12})$$

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 (B12) 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  $\alpha_{(n)} \lesssim \widehat{\alpha}_{(n)}$  and  $\beta_{(n)} \lesssim \widehat{\beta}_{(n)}$  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 (B12), we can calculate the difference  $\Omega_{\text{PDD}}^{(2)}(\widehat{\Omega} + \delta\Omega) - \Omega_{\text{PDD}}^{(2)}(\widehat{\Omega})$  using the expression for  $\Omega_{\text{PDD}}^{(2)}$  from Eq. (14). In the end,

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

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

Having showed its faithfulness, we are now confident in using the estimator Eq. (B2) to study the full  $\Omega_{(n)}$  of CDD $_n$ . In particular, we have  $\alpha_{(n)} \approx \widehat{\alpha}_{(n)} = 4^n \alpha$  and  $\beta_{(n)} \approx \widehat{\beta}_{(n)}$ . Similar as the PDD case, we can derive a strict bound and an approximate bound. For the strict bound, we have

$$\begin{aligned}\widehat{\beta}_{(n)} &\leq 2^{n(n+1)} \|\text{ad}_{B_0}^n(B_1)\| \\ &+ 2^{n^2} \|\text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\})\| \\ &\leq 2^{n(n+1)} \alpha^{n-1} \beta (2^n \alpha + \alpha + \beta).\end{aligned}\quad (\text{B13})$$



- 
- [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).