

# Efficacy of concatenated dynamical decoupling sequences

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, CNRS-UCA-SU-NUS-NTU International Joint Research Unit, Singapore

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) through logical codes [12–18]. Compared with QEC, dynamical decoupling is typically more economical as it require no encoding of logical qubits nor real-time close-loop control through constant syndrome measurement. All that needed is to put the system on a hook of pulses. In practice, one can choose to only implement DD or to incorporate DD within a standard QEC scheme as a “layer of defense” against the noise [19].

The use of DD does not, however, come at no cost: To implement it, one has to apply multiple pulses, which can be imperfect, and possibly even add noise to the quantum system. 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.

## II. PRELIMINARIES

### A. Channel Distance and Channel Infidelity

Trace distance and fidelity are two common ways to compare quantum states and have also been employed to study quantum channels. To distinguish two quantum states, we can calculate their trace distance by  $D(\rho, \sigma) \equiv \frac{1}{2}\|\rho - \sigma\|_1$  and fidelity by  $F(\rho, \sigma) \equiv \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ . The norm subscript in this paper follows the Schatten norm class,  $\|A\|_p \equiv \text{Tr}(|A|^p)^{1/p}$ , where  $p = 1, 2, \infty$  is also recognized as the trace norm, the Frobenius norm and the spectral norm respectively. The norm without explicit subscript is understood as the spectral norm.

For a quantum channel  $\mathcal{E}$ , the input-state-dependent channel distance and channel fidelity are defined by comparing the input state  $\rho$  with the output state  $\mathcal{E}(\rho)$ . We further strict the input state to be pure, where the fidelity expression can be simplified as the Hilbert-Schmidt inner product. We denote the identity channel by  $\mathcal{I}$  and define infidelity as  $\text{InF} \equiv 1 - F$ . It follows that

$$D(\mathcal{E}|\rho) \equiv D(\rho, \mathcal{E}(\rho)) = \frac{1}{2}\|(\mathcal{I} - \mathcal{E})(\rho)\|_1, \quad (1)$$

$$\text{InF}(\mathcal{E}|\rho) \equiv 1 - F(\rho, \mathcal{E}(\rho)) = \text{tr}(\rho(\mathcal{I} - \mathcal{E})(\rho)). \quad (2)$$

According to the Fuchs-van de Gaaf inequality, the channel distance and infidelity are mutually bounded by:

$$\text{InF} \leq D \leq \sqrt{\text{InF}}, \quad (3)$$

for any fixed pure input state  $\rho$  and quantum channel  $\mathcal{E}$ .

To get rid of the state-dependence in the channel distance and channel infidelity, we take the maximization over all possible pure input state. Since the set of all density operators  $\mathcal{D}(\mathcal{H})$  is compact and convex, for any Hermitian-preserving superoperator  $\mathcal{E}$ , the optimal fidelity value is always attained on a pure state. These state-independent quantities can be written as:

$$D_{\max}(\mathcal{E}) = \max_{\rho \in \mathcal{D}(\mathcal{H})} \frac{1}{2}\|(\mathcal{I} - \mathcal{E})(\rho)\|_1 \equiv \frac{1}{2}\|\mathcal{I} - \mathcal{E}\|_1, \quad (4)$$

$$\text{InF}_{\max}(\mathcal{E}) \equiv \max_{\rho \in \mathcal{D}(\mathcal{H})} \text{tr}(\rho(\mathcal{I} - \mathcal{E})(\rho)). \quad (5)$$

The inequalities (3) still hold for these state-independent measures.

## B. Error Phase

Channel distance and channel fidelity are dynamics-agnostic measures of the noise strength. If we restrict the quantum channels to the set of completely-positive and trace-preserving (CPTP) maps, which are channels preserving the set of physical states even on arbitrarily extended Hilbert space, we can define an additional performance indicator—the error phase [5].

A generic CPTP map can be composed by applying a unitary evolution operator on an initially uncoupled state of the system and bath, followed by tracing out the bath. For the unitary evolution, we assign a Hermitian generator  $\Omega \in \mathfrak{h}_{SB}$  from the Lie algebra of the combined Hilbert space  $\mathcal{H}_S \otimes \mathcal{H}_B$ . The output system state, given an arbitrary initial state  $\rho_S \in \mathcal{H}_S$ , can be written as

$$\mathcal{E}(\rho_S) = \text{tr}_B (e^{-i\Omega} (\rho_S \otimes \rho_B) e^{i\Omega}), \quad (6)$$

where  $\rho_B \in \mathcal{H}_B$  is the initial bath state;  $\text{tr}_B(\cdot)$  is the partial trace over  $\mathcal{H}_B$ . Since the noise map  $\mathcal{E}$  is completely determined by  $\Omega$  and  $\rho_B$ , calculating the channel distance or channel fidelity requires full knowledge of them. This could be impractical sometimes and we want a easier quantity to work with.

One possibility is to instead consider the norm  $\|\Omega\|$ . However, a moment's thought reveals that this measure is flawed: one is allowed to add any constant shift or any irrelevant ancilla state interaction to the generator without changing the channel, yet both will change the norm. To address this issue, we set the generator to be traceless  $\text{tr}(\Omega) = 0$ , and split it into two parts:

$$\Omega = \Omega_B + \Omega_{SB}, \quad (7)$$

this can be formally achieved by defining a pair of projection operators:

$$\mathcal{P}_0(\cdot) = \frac{I_S}{d_S} \otimes \text{trs}(\cdot), \quad \mathcal{P}_c = \mathbb{I} - \mathcal{P}_0. \quad (8)$$

where  $I_S$  is the identity operator on the system and  $d_S$  is the bath dimension. We recognize the “pure bath” part  $\Omega_B \equiv \mathcal{P}_0(\Omega)$  and the coupling part  $\Omega_{SB} \equiv \mathcal{P}_c(\Omega)$ . These two parts contribute differently to the noise strength. Intuitively,  $\Omega_{SB}$  couples the system with the bath and should contribute directly to the noise; while  $\Omega_B$  is exclusive to the bath and should not have much impact. Following this simple reasoning, we define the error phase as the norm of the coupling part, that is,  $\|\Omega_{SB}\|$ . For a sanity check, suppose an ancilla system  $\mathcal{H}_A$  completely unrelated to  $\mathcal{H}_S \otimes \mathcal{H}_B$  is taken into account. Consequently, the generator is modified to  $\Omega' = \Omega_B \otimes \Omega_A + \Omega_{SB} \otimes I_A$ , whose error phase  $\|\Omega_{SB} \otimes I_A\| = \|\Omega_{SB}\| \|I_A\| = \|\Omega_{SB}\|$  remains unaffected.

In general, for any Lie algebra element  $\Omega \in \mathfrak{h}_{SB}$ , we can recognize the splitting according to Eq. (7). To trace the magnitude of both the pure bath part and coupling part, we introduce the “two-component norm” as a map

from the Lie algebra to a pair of positive numbers:

$$\|\Omega\| \equiv (\|\Omega_B\|, \|\Omega_{SB}\|). \quad (9)$$

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

$$\|(U_S \otimes U_B)\Omega(U_S^\dagger \otimes U_B)\| = \|\Omega\|, \forall U_S(U_B) \in \mathbf{U}(d_{S(B)}), \quad (10)$$

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

## C. Error phase vs. Channel infidelity

To capture more precisely how the error phase relate to the noise strength of a CPTP map, we establish its connection to the channel infidelity. Such connection rely on the assumption that the noise is weak, in the sense that Hermitian operator  $\Omega$  in Eq. (D5) associated with the noise map is small enough.

Invoking the Baker-Hausdorff lemma [20], we expand the output state in Eq. (D5) as a series in the powers of  $\Omega$ . With a simple rearrange,

$$(\mathcal{I} - \mathcal{E})(\rho_S) = - \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \text{tr}_B \text{ad}_\Omega^n(\rho_S \otimes \rho_B), \quad (11)$$

where the adjoint operator is defined by the commutator  $\text{ad}_\Omega(\cdot) \equiv [\Omega, \cdot]$ , with  $\text{ad}_\Omega^n$  involving  $n$ -fold commutation with  $\Omega$ . According to definition Eq. (D8), we can express the channel infidelity as a series as well by taking trace product with  $\rho_S$  for each term in Eq. (11),

$$\text{InF}(\mathcal{E}|\rho_S) \equiv \sum_{n=1}^{\infty} \text{InF}^{(n)}(\mathcal{E}|\rho_S). \quad (12)$$

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

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

$$\text{tr}(\rho_S \text{tr}_B \text{ad}_\Omega(\rho_S \otimes \rho_B)) = \text{tr}(\rho_S [\text{tr}_B(\Omega \rho_B), \rho_S]) = 0. \quad (13)$$

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

$$\text{tr}_B \text{ad}_\Omega = 0, \quad \forall \Omega_B \in \mathfrak{h}_B. \quad (14)$$

The proof is straightforward by applying the above operator to a generic linear operator  $\sum_\alpha S_\alpha \otimes B_\alpha$  on the

combined Hilbert space. We now substitute the Eq. (7) into the double commutator term  $\sim \text{ad}_\Omega^2$  in Eq. (11) and expand. After applying the Jacobi identity— $\text{ad}_A \text{ad}_B = \text{ad}_B \text{ad}_A + \text{ad}_{[A,B]}$ —and invoking Eq. (13) and Eq. (14), we have

$$\text{tr}(\rho_S \text{tr}_B \text{ad}_\Omega^2 \rho_S \otimes \rho_B) = \text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega_{SB}}^2 \rho_S \otimes \rho_B). \quad (15)$$

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

$$\begin{aligned} 0 \leq \text{InF}^{(2)}(\mathcal{E}|\rho_S) &= \frac{1}{2} \text{tr}(\rho_S \text{tr}_B \text{ad}_{\Omega_{SB}}^2 (\rho_S \otimes \rho_B)) \\ &\leq \frac{1}{2} \|\rho_S \otimes I_B\|_\infty \|\text{ad}_{\Omega_{SB}}^2 (\rho_S \otimes \rho_B)\|_1 \\ &\leq 2\|\Omega_{SB}\|_\infty^2 \|\rho_S \otimes \rho_B\|_1 \leq 2\|\Omega_{SB}\|_\infty^2, \end{aligned} \quad (16)$$

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

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

On the other hand, the error phase provides only a rough indicator of the channel noise strength and could go wrong if special symmetries present. The bounds in Eq. (16) hints the problem: the error phase is good at measuring the upper bound but fails at the lower bound. In fact, one can think of an extreme example: in some symmetry-protected subspace where  $[H, \rho_S \otimes \rho_B] = 0$ , the whole evolution is noiseless, thus  $\text{InF}(\mathcal{E}|\rho_S) = 0$ , but the error phase could be arbitrary. In some other cases where only  $\text{InF}^{(2)}(\mathcal{E}|\rho_S) = 0$  and higher order terms persist, the infidelity is no longer approximated by  $\|\Omega_{SB}\|^2$ .

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

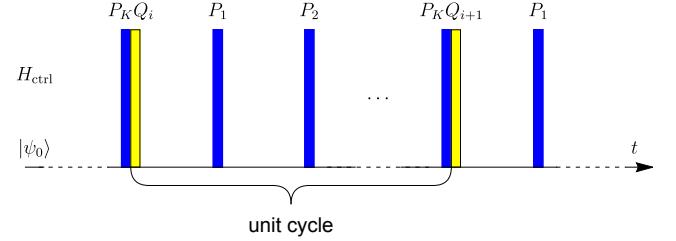
$$D_{\max}(\alpha, \beta) = \max_{\|\Omega\|=(\alpha, \beta)} D_{\max}(\mathcal{E}), \quad (17)$$

$$\text{InF}_{\max}(\alpha, \beta) = \max_{\|\Omega\|=(\alpha, \beta)} \text{InF}_{\max}(\mathcal{E}), \quad (18)$$

where the  $D_{\max}(\mathcal{E})$  and  $\text{InF}_{\max}(\mathcal{E})$  are defined in Eq. (4) and Eq. (5) for  $\mathcal{E}$  generated by  $\Omega$ . This allows us to establish a fine-grained hierarchy of functions to gauge performance. For channel distance,  $D_{\min}(\alpha, \beta) \leq D_{\min}(\mathcal{E}) \leq D(\mathcal{E}|\rho_S) \leq D_{\max}(\mathcal{E}) \leq D_{\max}(\alpha, \beta)$ ; and for the channel  $\text{InF}_{\min}(\alpha, \beta) \leq \text{InF}_{\min}(\mathcal{E}) \leq \text{InF}(\mathcal{E}|\rho_S) \leq \text{InF}_{\max}(\mathcal{E}) \leq \text{InF}_{\max}(\alpha, \beta)$ , where the functions with “min” subscript are defined by replacing the corresponding maximization with minimization. Through the  $[H, \rho(0)] = 0$  example we know that  $D_{\min}(\alpha, \beta) = \text{InF}_{\min}(\alpha, \beta) = 0$ .

#### D. Dynamical decoupling

To preserve the memory or computation for a given length of time, a fixed sequence of quantum gates are applied periodically to the central system. We call the periodic applied sequence the unit cycle. Let  $K$  be the number of gates in a cycle be and the  $i$ -th gate in a unit cycle denoted as  $P_i$ . Then the full sequence of gates applied to the system can be schematically represented as follows



In the diagram,  $Q_i$  is the computational gate ( $= I$  for quantum memory) to be “protected” with dynamical decoupling; the dash represents the “wait time” between two gates; two gates without dash in between should be combined into a single gate. Given its periodic structure, to study the effect of the DD sequence, we only need to focus on the time evolution of the unit cycle.

All DD schemes share some fundamental similarities. First, the gates are always chosen from a specific group of transformations  $P_i \in \mathbb{G}$ ,  $1 \leq i \leq K$ . Second, all sequences must satisfy the constraint that they multiply to identity:

$$P_K P_{K-1} \cdots P_1 = I, \quad (19)$$

so that we do not introduce additional transformations to the system. On the other hand, different strategies could differ in the following aspects. ① The transformation group  $\mathbb{G}$ . For single qubit, the most obvious choice is perhaps the Pauli group  $\{I, X, Y, Z\}$ . Some simple schemes such as the spin-echo [21] and the CPMG sequence [22, 23] use only the subgroup  $\{I, Z\}$ . ② The cycle length  $K$  and the specific arrangement of the gates. ③ The time at which the gates are applied. In particular, we have constant gate-interval schemes, such as the periodic DD (PDD) [1] and the concatenated DD (CDD)

[4]; and variable gate-interval schemes, such as the UDD [7], the nested UDD (NUDD) [9] and the quadratic DD (QDD) [11]. In this paper, we focus specifically on the constant time-interval schemes with Pauli gates. The underlying physical model is a constant-frequency quantum circuit, whose operating frequency is limited by a finite switch-time  $\tau_0 > 0$  between consecutive gates. A DD unit cycle of length  $K$  is assumed to start from time  $t = 0$  and end at  $t = K\tau_0 \equiv T$ .

Two potential sources of error associated with the DD procedure can be identified: the noise induced by interaction between the system and the bath, which we aim to eliminate; and the error introduced by the imperfect DD gates. Here we refer to the case where the second noise source is absent as the idealistic decoupling; and the model where both sources of noise are present as the noisy decoupling.

### III. IDEALISTIC DECOUPLING

#### A. Generic theory

We model the noise associated with the free evolution by CPTP maps. For the time evolution up to  $\tau_0$ , we can find a Hamiltonian operator  $H$  responsible for generating the unitary transformation between the system and bath,

$$U_f(\tau_0) = e^{-i\tau_0 H} \equiv e^{-i\Omega}, \quad (20)$$

where  $\Omega = \tau_0 H$  is also referred to as the dimensionless Hamiltonian, and generates the free-evolution noise channel according to Eq. (D5).

Ideally, the DD gates  $\{P_i\}$  can be modeled as perfect pulses applied at time instances  $\{t_i = i\tau_0, i = 1, \dots, K\}$ . We make the assumption that the Hamiltonian remains constant within the duration of the DD cycle, the time evolution operator for the cycle can be written as

$$U_{DD} = P_K e^{-i\Omega} \cdots P_2 e^{-i\Omega} P_1 e^{-i\Omega} = e^{-i\Omega_{DD}}, \quad (21)$$

where we have assigned an effective Hamiltonian  $\Omega_{DD}$  for the whole cycle.

The constraint Eq. (19) on the DD gates, together with the fact that all gates form a group, allow us to split each gate into a product of two hypothetical gates,  $P_i = G_{i+1}^\dagger G_i$  ( $1 \leq i \leq K$ ) with  $G_1 = G_{K+1} = I$ . In light of this, specifying a sequence of DD gates  $(P_1, \dots, P_K)$  is equivalent to specifying a sequence of gates  $(G_2, \dots, G_K)$  without any constraint. Substituting the splitting into Eq. (21) and taking the matrix logarithm, we get

$$\Omega_{DD} = i \log(e^{-i\Omega_K} \cdots e^{-i\Omega_2} e^{-i\Omega_1}), \quad (22)$$

where we define  $\Omega_i \equiv G_i \Omega G_i^\dagger$ . The necessary convergence criteria is also assumed to ensure the existence of the matrix logarithm.

The Eq. (22) gives the formal solution of the DD transformation on the noise dynamics, but in general cannot be analytically solved. To proceed, one need to resort

to approximation methods. This can be done by combining the matrix exponentials via the BCH formula, or by performing the Magnus expansion for time-dependent Hamiltonian. Both approaches give the identical result by expressing the exponent as an infinite series,

$$\Omega_{DD} = \sum_{m=1}^{\infty} \Omega_{DD}^{(m)}, \quad (23)$$

where the size of the  $k$ th term is of the order  $\mathcal{O}(\|\Omega\|^k)$ . Compared with the BCH formula approach, the Magnus expansion approach allows easier derivation for large  $K$  and predicts a broader convergence bound. From the theory of Magnus expansion [24], a sufficient condition for (23) to converge by absolute value is

$$K\|\Omega\|_\infty < \xi = 1.0868 \dots \quad (24)$$

The absolute-converge criterion is important for any finite truncation of the series (23) to make a reasonable approximation. In the following analysis, we assume such criterion is met.

Following the Magnus series formula, terms in Eq. (23) can be derived. In particular, the first three terms can be calculated by the formula:

$$\Omega_{DD}^{(1)} = \sum_i \Omega_i, \quad \Omega_{DD}^{(2)} = -\frac{i}{2} \sum_{i>j} [\Omega_i, \Omega_j], \quad (25)$$

and

$$\Omega_{DD}^{(3)} = -\frac{1}{6} \sum_{i\geq j\geq k} \frac{[\Omega_i, [\Omega_i, \Omega_k]] + [\Omega_k, [\Omega_j, \Omega_i]]}{\text{sym}(i, j, k)}, \quad (26)$$

where the symmetry factor  $\text{sym}(i, j, k)$  is defined as the factorial product of the degeneracy of the indices, e.g.,  $\text{sym}(3, 2, 2) = 1! \times 2! = 2$ . In general, the  $m$ th order term can be written as a combination of nested commutators in  $m$ -fold copies of  $\Omega_i$ . One can in principle apply the Magnus formula to obtain higher order terms, but the expressions get complicated rapidly and effectively make direct derivations impractical. Instead, clever tricks that take advantage of the sequence structure need to be used.

According to Eq. (22), the DD protocol can be regarded as a transformation  $\Omega \rightarrow \Omega_{DD}$ . Since the noise map  $\mathcal{E}$  is a function of  $\Omega$  and  $\rho_B$ , the transformation on  $\Omega$  takes the original quantum channel into a new one  $\mathcal{E} \rightarrow \mathcal{E}_{DD}$ . As is argued in Sec. II C, 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

$$\|\Omega\| \equiv (\alpha, \beta) \quad \text{and} \quad \|\Omega_{DD}\| \equiv (\alpha_{DD}, \beta_{DD}). \quad (27)$$

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  $\beta_{DD}/\beta < 1$ , 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 (23) have vanishing coupling part:

$$\mathcal{P}_c(\Omega_{\text{DD}}^{(k)}) = 0, \text{ for } 1 \leq k \leq n. \quad (28)$$

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

$$\beta_{\text{DD}} = \left\| \sum_{k=n+1}^{\infty} \mathcal{P}_c(\Omega_{\text{DD}}^{(k)}) \right\| \simeq (\alpha, \beta)^{n+1}, \quad (29)$$

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. (29) 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.

Meanwhile we may also employ the more sophisticated gauge of the channel distance and the channel infidelity to characterize DD performance. These indicators could accommodate the weakness in using error phase, particularly in cases with special symmetries. For example, if one prepares the system in some decoherence-free subspace, applying the DD transformation  $\Omega \rightarrow \Omega_{\text{DD}}$  would break the initial symmetries and therefore increase the noise level according to the fidelity or distance measures. But this argument does not speak against the use of DD in general. The only promise of DD is to “decouple” the interaction Hamiltonian by reducing the magnitude of the coupling strength, and we shall consider the objective accomplished if DD serves such purpose.

In the following, we study two notable DD schemes with fixed pulsed interval and Pauli group pulses: periodic DD (PDD) [1] and concatenated DD (CDD) [5]. The PDD scheme can be seen as a special case of CDD. Our analysis showcases the steps to derive the fault-tolerance properties of DD and can be generalized to study other schemes.

## B. PDD

For a single-qubit system coupled with a bath, we denote the free evolution Hamiltonian for time  $\tau_0$  as:

$$\Omega = \tau_0 H = \sigma_0 \otimes B_0 + \sum_{i=1}^3 \sigma_i \otimes B_i, \quad (30)$$

where  $\{\sigma_i\}$  are the system space Pauli operators and  $\{B_i\}$  are the Hermitian bath operators with  $\text{tr}(B_0) = 0$ . The

bath operators  $B_i$  can be arbitrary dimensional, but are normalized such that

$$\|\Omega\| = \left( \left\| \sum_{i=1}^3 \sigma_i \otimes B_i \right\| \right) \equiv \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (31)$$

To reduce noise, we perform dynamical decoupling by periodically applying the pulse cycle “ $-X - Z - X - Z$ ”. This is also known as PDD scheme. This cycle is sometimes referred to as the universal decoupling sequence and it can suppress all types of noise on the qubit. Under the ideal-pulse assumption, the time evolution operator for the PDD cycle is:

$$Z e^{-i\Omega} X e^{-i\Omega} Z e^{-i\Omega} X e^{-i\Omega} \equiv e^{-i\Omega_P}, \quad (32)$$

where an effective Hamiltonian  $\Omega_P$  is defined for the time evolution. The transformation  $\Omega \rightarrow \Omega_P$  can be thought as an operator-valued function  $\Omega_P(\Omega)$ . Furthermore, since the resulting  $\Omega_P$  can be expressed as the Magnus series, each term in the series can be regarded as a function:

$$\Omega_P = \Omega_P(\Omega) = \sum_{m=1}^{\infty} \Omega_P^{(m)}(\Omega). \quad (33)$$

Applying the formula (25), the first two terms in the series can be computed as:

$$\Omega_P^{(1)}(\Omega) = \sigma_0 \otimes 4B_0 \quad (34)$$

$$\begin{aligned} \Omega_P^{(2)}(\Omega) &= \sigma_1 \otimes (-4i[B_0, B_1]) \\ &\quad + \sigma_2 \otimes (-2i[B_0, B_2] - 2\{B_1, B_3\}). \end{aligned} \quad (35)$$

This explicitly verifies that PDD achieves first order decoupling as the coupling part vanishes for  $\Omega_P^{(1)}$ .

## C. CDD

Based upon PDD, a class of pulse schemes is proposed by Khodjasteh and Lidar [4] and named as CDD. In this section, we present an independent analysis on CDD, with an emphasis on the fault-tolerance (FT) perspective. Previous paper on fault-tolerance also exist [5] with a focus on the quantum memory model. Our analysis is leant towards the quantum computation scenario. This results in different conclusions regarding the FT performance of CDD. Steps used in the derivation of CDD here also facilitate the development of the threshold criterion with noisy pulses.

To define the CDD scheme, we first denote the free time evolution operator as  $U_{(0)} \equiv e^{-i\Omega}$ , and the PDD time evolution operator as  $U_{(1)} \equiv e^{-i\Omega_P}$ . Higher level time evolution operators are defined recursively through the application of the PDD pulses on lower level maps:

$$U_{(k+1)} \equiv Z U_{(k)} X U_{(k)} Z U_{(k)} X U_{(k)}. \quad (36)$$

Such recursion goes on until a predetermined maximum level is reached. For concatenation level  $n$ , the resulting

pulse sequence is denoted as  $\text{CDD}_n$ . We also associate an effective Hamiltonian for the time evolution operator at each level,  $U_{(k)} \equiv e^{-i\Omega_{(k)}}$ ,  $k = 1, \dots, n$ . To construct the actual gate sequence for  $\text{CDD}_n$ , we start with the recursive definition and then flatten out the nested levels by combining consecutive gates. This leads to the time evolution operator  $U_{(n)}$  of the form (21) with  $K = 4^n$  gates to apply.

Take  $\text{CDD}_2$  as an example, the eventual flattened gate sequence would be “ $-X - Z - X - Y - X - Z - X - I - X - Z - X - Y - X - Z - X - I$ ”. And in terms of the effective Hamiltonian, we have the following map:

$$\begin{aligned}\Omega_{(2)} = i \log(e^{-i\Omega} e^{-iX\Omega X} e^{-iY\Omega Y} e^{-iZ\Omega Z} e^{-iX\Omega X} \\ e^{-i\Omega} e^{-iZ\Omega Z} e^{-iY\Omega Y} e^{-iY\Omega Y} e^{-iZ\Omega Z} e^{-i\Omega} \\ e^{-iX\Omega X} e^{-iZ\Omega Z} e^{-iY\Omega Y} e^{-iX\Omega X} e^{-i\Omega}).\end{aligned}$$

After applying Eq. (25), and Eq. (26) to calculate the first three Magnus series, we find  $\Omega_{(2)}^{(1)} = \sigma_0 \otimes 16B_0$ ,  $\Omega_{(2)}^{(2)} = 0$ , and  $\Omega_{(2)}^{(3)} \neq 0$  (for both pure-bath part and coupling part). This shows that  $\text{CDD}_2$  achieves second order decoupling.

It was suggested by the authors of CDD that the decoupling order for  $\text{CDD}_n$  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  $\text{CDD}_n$ , it is infeasible to flatten out the gates and calculate the combined Magnus series directly. Instead it is easier to study the iteration map

$$\Omega_{(n)} = \Omega_P(\Omega_{(n-1)}) = \sum_{m=1}^{\infty} \Omega_P^{(m)}(\Omega_{(n-1)}), \quad (37)$$

where the update map  $\Omega_P$  is defined in Eq. (33). Backtracking the map from level  $n$  to level 0, we have  $\Omega_{(n)} = \Omega_P \circ \dots \circ \Omega_P(\Omega) \equiv (\Omega_P)^n(\Omega)$ . In the end, the effective Hamiltonian  $\Omega_{(n)}$  can be expanded as:

$$\Omega_{(n)} = \sum_{m=1}^{\infty} \Omega_{(n)}^{(m)}(\Omega), \quad (38)$$

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. (37), where each  $\Omega_P^{(m)}$  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  $\widehat{\Omega}_{(n)}$  which is easier to solve. Formally we have an estimator-error pair  $(\widehat{\Omega}_{(n)}, \delta\Omega_{(n)})$  such that

$$\Omega_{(n)} = \widehat{\Omega}_{(n)} + \delta\Omega_{(n)}. \quad (39)$$

The estimator  $\widehat{\Omega}_{(n)}$  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 (38). 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  $\widehat{\Omega}_{(n)}$ , and demanding

$$\|\delta\Omega_{(n)}\| \equiv \begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \ll \|\widehat{\Omega}_{(n)}\| \equiv \begin{pmatrix} \widehat{\alpha}_{(n)} \\ \widehat{\beta}_{(n)} \end{pmatrix}, \quad (40)$$

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 C for more rigorous descriptions on ordering polynomials by their leading powers. Furthermore, after taking the two-component norm and applying triangular inequality for Eq. (39), we find,

$$\widehat{\beta}_{(n)} - \delta\beta_{(n)} \leq \beta_{(n)} \leq \widehat{\beta}_{(n)} + \delta\beta_{(n)}. \quad (41)$$

According to Eq. (29), 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  $\widehat{\beta}_{(n)}$  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:

$$\widehat{\Omega}_{(n)} \equiv (\Omega_P^{(1)} + \Omega_P^{(2)})(\widehat{\Omega}_{(n-1)}) = (\Omega_P^{(1)} + \Omega_P^{(2)})^n(\Omega). \quad (42)$$

With  $\Omega_P^{(1)}, \Omega_P^{(2)}$  explicitly defined in Eq. (34) and Eq. (35), and the starting point  $\widehat{\Omega}_{(0)} \equiv \Omega$ , the estimator can shown to be specified by the formula:

$$\begin{aligned}\widehat{\Omega}_{(n)} = & \sigma_0 \otimes 4^n B_0 \\ & + \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B_0}^n(B_1) \\ & + \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\}).\end{aligned} \quad (43)$$

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

$$\begin{pmatrix} \widehat{\alpha}_{(n)} \\ \widehat{\beta}_{(n)} \end{pmatrix} \simeq \begin{pmatrix} \alpha \\ \alpha^n \beta + \alpha^{n-1} \beta^2 \end{pmatrix}, \quad (44)$$

where  $\simeq$  represents equal the leading powers as described in Appendix C. 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:

$$\begin{pmatrix} \delta\alpha_{(n)} \\ \delta\beta_{(n)} \end{pmatrix} \lesssim \begin{pmatrix} \tau_0^3 \\ \tau_0^{n+2} \end{pmatrix}. \quad (45)$$

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:

$$\|\delta\Omega_{(1)}\| \leq \sum_{m=3}^{\infty} \|\Omega_{(1)}^{(m)}\| \simeq \|\Omega_{(1)}^{(3)}\| \lesssim \left(\frac{\tau_0^3}{\tau_0^3}\right). \quad (46)$$

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

$$\begin{aligned} \delta\Omega_{(n+1)} &= \Omega_P(\Omega_{(n)}) - (\Omega_P^{(1)} + \Omega_P^{(2)})(\widehat{\Omega}_{(n)}) \\ &= \Omega_P^{(1)}(\delta\Omega_{(n)}) + \sum_{m=2}^{\infty} \Omega_P^{(m)}(\Omega_{(n)}) - \Omega_P^{(2)}(\widehat{\Omega}_{(n)}), \end{aligned} \quad (47)$$

where we have used the fact that  $\Omega_P^{(1)}$  is a linear map. Assuming Eq. (45) 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

$$\left\| \sum_{m=3}^{\infty} \Omega_P^{(m)}(\widehat{\Omega}_{(n)}) \right\| \ll \|\widehat{\Omega}_{(n+1)}\|.$$

If we take  $\delta\Omega_{(n)} = 0$ , Eq. (47) would identify 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. (47) 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. (47):

$$\begin{aligned} \|\delta\Omega_{(n+1)}\| &\lesssim \|\Omega_P^{(1)}(\delta\Omega_{(n)})\| + \|\Omega_P^{(3)}(\Omega_{(n)})\| \\ &\quad + \|\Omega_P^{(2)}(\widehat{\Omega}_{(n)} + \delta\Omega_{(n)}) - \Omega_P^{(2)}(\widehat{\Omega}_{(n)})\|, \end{aligned} \quad (48)$$

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 (48) is easy,

$$\|\Omega_P^{(1)}(\delta\Omega_{(n)})\| = \left(\frac{4\delta\alpha_{(n)}}{0}\right) \simeq \left(\frac{\tau_0^3}{0}\right).$$

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

$$\|\Omega_P^{(3)}(\Omega_{(n)})\| \lesssim \left(\frac{\beta_{(n)}^2(\alpha_{(n)} + \beta_{(n)})}{\beta_{(n)}(\alpha_{(n)} + \beta_{(n)})^2}\right) \simeq \left(\frac{\tau_0^{2n+2}}{\tau_0^{n+3}}\right),$$

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. (26). To bound the third term in (48), we can calculate the difference  $\Omega_P^{(2)}(\widehat{\Omega} + \delta\Omega) - \Omega_P^{(2)}(\widehat{\Omega})$  using the expression for  $\Omega_P^{(2)}$  from Eq. (35). In the end,

$$\begin{aligned} \|\Omega_P^{(2)}(\widehat{\Omega} + \delta\Omega) - \Omega_P^{(2)}(\widehat{\Omega})\| &\lesssim \left(\frac{0}{\delta\alpha_{(n)}\delta\beta_{(n)}}\right) \\ &\quad + \left(\frac{0}{\widehat{\beta}_{(n)}\delta\alpha_{(n)} + \widehat{\alpha}_{(n)}\delta\beta_{(n)} + \widehat{\beta}_{(n)}\delta\beta_{(n)}}\right) \simeq \left(\frac{0}{\tau_0^{n+3}}\right). \end{aligned}$$

Since all three terms are bounded by  $(\tau_0^3 \tau_0^{n+3})^T$ , so is their sum. This proves our assertion Eq. (45).  $\square$

Having showed its faithfulness, we are now confident in using the estimator Eq. (38) to study the full  $\Omega_{(n)}$  of CDD <sub>$n$</sub> . 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

$$\begin{aligned} \widehat{\beta}_{(n)} &\leq 2^{n(n+1)} \|\text{ad}_{B_0}^n(B_1)\| \\ &\quad + 2^{n^2} \|\text{ad}_{B_0}^{n-1}([B_0, B_2] - i\{B_1, B_3\})\| \\ &\leq 2^{n(n+1)} \alpha^{n-1} \beta (2^n \alpha + \alpha + \beta). \end{aligned} \quad (49)$$

From this upper bound, we see that as the concatenation level goes up,  $\widehat{\beta}_{(n)}$  would first decrease due to the small  $\alpha$  value, but then start to increase due to the super-exponentially growing coefficient. This suggests the existence of a maximally allowed concatenation level, beyond which increasing the concatenation level would only introduce more noise. This can be seen explicitly from Fig. 1, where the relevant sizes of  $\widehat{\alpha}_{(n)}$  and  $\widehat{\beta}_{(n)}$  are plotted against the concatenation level in the logarithmic scale. For higher concatenation level to make sense,

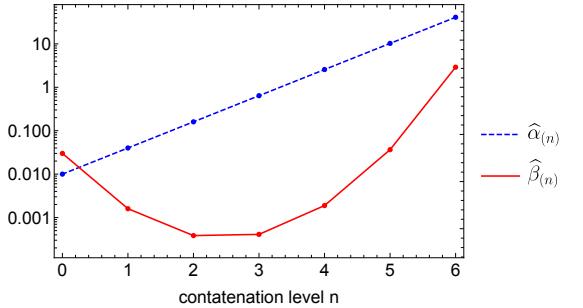


FIG. 1: The estimator size for  $\widehat{\alpha}_n$  and the error phase  $\widehat{\beta}_n$  as a function of concatenation level  $n$ . In the graph, the bare Hamiltonian is set to take on the value  $\alpha = 0.01$  and  $\beta = 0.01$ .

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. (49) of the CDD<sub>n</sub> to that of the CDD<sub>n+1</sub>,

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

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:

$$n_{\max} \leq -\log_4 \alpha - \frac{3}{2}. \quad (50)$$

The existence of a maximal concatenation level does not contradict with the higher order decoupling achieved by higher level CDD. It simply means that the correspondence between higher decoupling order and better noise suppression is not always true. Generally, if the absolute converge criteria Eq. (24), is fulfilled, a higher  $n$  means the leading order term is smaller. But if the DD scheme is designed such at the sequence length  $K$  grows exponentially with  $n$ , increasing  $n$  will break the absolute convergence and the decoupling order is no longer a reasonable indicator.

#### IV. NOISY DECOUPLING AND FAULT-TOLERANCE

To determine whether DD should be implemented on a particular system, one need to take on the fault tolerance perspective by weighing the costs and benefits of DD. If the benefits of doing DD outweigh the costs, DD should be encouraged; otherwise DD should be avoided.

To begin with, it is important to distinguish whether the task is to preserve states in quantum memory, or to protect gates from noise in quantum computation. For quantum memory, the goal is to preserve quantum states for as long as possible. Whether implementing DD or not will not change the total evolution time. For quantum computation, on the other hand, the introduction of extra DD pulses dilate the time interval between computational gates. Longer gate interval allows more time for noise to develop, potentially counteracting the benefit of DD. Therefore despite the same transformation  $\Omega \rightarrow \Omega_{\text{DD}}$  for both cases, the relevant transformation in terms of the time evolution operator, is  $e^{-iK\Omega} \rightarrow e^{-i\Omega_{\text{DD}}}$  for quantum memory and  $e^{-i\Omega} \rightarrow e^{-i\Omega_{\text{DD}}}$  for quantum computation. In terms of the error phase, DD introduces the transformation:  $\beta \rightarrow \beta_{\text{DD}}$ . For a DD sequence with cycle length  $K$ , we can formulate the criteria for noise reduction in terms of error phase as

$$\begin{aligned} \beta_{\text{DD}} &\leq K\beta && \text{for quantum memory,} \\ \beta_{\text{DD}} &\leq \beta && \text{for quantum computation.} \end{aligned}$$

In this the paper, we focus primarily on the quantum computation scenario.

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 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. idealistic

To determine when is PDD useful, we compare the error phase by solving  $\beta_{\text{PDD}} \leq \beta$ . A precise solution is in general impossible as the full series (33) is unknown. We instead look for approximate solution using the leading order approximation and solve for

$$\beta_{\text{PDD}} \approx \|\Omega_{\text{P}}^{(2)}(\Omega)\| \leq \beta \quad (51)$$

To solve this inequality, we need to express the error phase in terms of the bath operators. We note the following inequalities:

$$\max_{i \in \{1, 2, 3\}} \|B_i\| \leq \left\| \sum_{i=1}^3 \sigma_i \otimes B_i \right\| \leq \sum_{i=1}^3 \|B_i\|. \quad (52)$$

The proof is given in Appendix B. We use these bounds to derive a “strict condition”. First,

$$\begin{aligned} \|\Omega_{\text{P}}^{(2)}(\Omega)\| &\leq 4\|[B_0, B_1]\| + 2\|i[B_0, B_2] + \{B_1, B_3\}\| \\ &\leq 8\alpha\|B_1\| + 4\alpha\|B_2\| + 4\|B_1\|\|B_3\| \\ &\leq 12\alpha\beta + 4\beta^2, \end{aligned} \quad (53)$$

we then bound the last line by  $\beta$  to arrive at:

$$12\alpha + 4\beta \leq 1. \quad (54)$$

On the other hand, the inequalities (52) inspires use to use the following approximation for the error phase:

$$\left\| \sum_{i=1}^3 \sigma_i \otimes B_i \right\| \approx \sqrt{\sum_{i=1}^3 \|B_i\|^2}. \quad (55)$$

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

$$\begin{aligned} \beta_{\text{PDD}}^2 &\approx \| -4i[B_0, B_1] \|^2 + \| -2i[B_0, B_2] - 2\{B_1, B_3\} \|^2 \\ &\leq (8\alpha\|B_1\|)^2 + (4\alpha\|B_2\| + 4\|B_1\|\|B_3\|)^2, \end{aligned} \quad (56)$$

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

$$\max_{x_1^2+x_2^2+x_3^2=1} (2\alpha x_1)^2 + (\alpha x_2 + \beta x_1 x_3)^2 \leq \frac{1}{16}. \quad (57)$$

Solution of this constrained optimization problem leads to:

$$\left( \frac{\beta}{\sqrt{3}} \leq \alpha \leq \frac{1}{8} \right) \text{ or } \left( 10\alpha^2 + \frac{9\alpha^4}{\beta^2} + \beta^2 \leq \frac{1}{4} \right). \quad (58)$$

This set of inequalities automatically implies (54), it gives a broader but less confident condition due to the approximation used for the error phase.

The criteria (54) and (58) 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 (24), the necessary condition here is  $4\|\Omega\|_\infty < \xi$ . Since  $\|\Omega\| \leq \alpha + \beta$ , we may use a more stringent convergence bound:

$$\alpha + \beta \leq \frac{1}{4}. \quad (59)$$

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. (D2), then simulate the PDD protocol and use matrix logarithm to find the actual  $\Omega_P$  and error phase  $\beta_{PDD}$ . For each fixed  $(\alpha, \beta)$  point, we numerically maximize the error phase reduction ratio  $\beta_{PDD}/\beta$  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 (54) and (58), together with the absolute convergence region (59) for comparison.

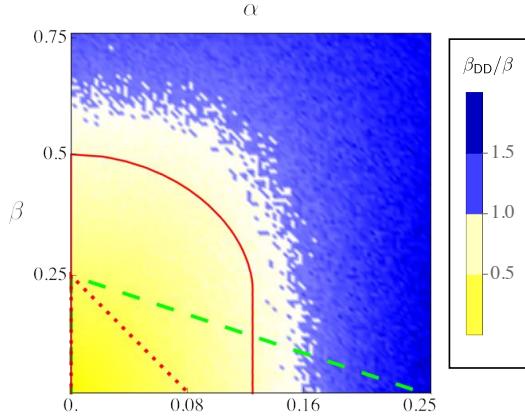


FIG. 2: The noise reduction ratio  $\beta_{DD}/\beta$  obtained through maximization over random bath operators plotted in the  $(\alpha, \beta)$  parameter plane. In comparison, the theoretical boundaries for noise reduction are also plotted, with dotted red line for (54) and solid red line for (58), also with the boundary for the absolute-convergence condition (59) in dashed green line. Note the ranges for  $\alpha$  and  $\beta$  are chosen differently for better visual effect.

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 (54) is too loose compared with the “approximate condition” (58), 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 (59), 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.

To best determine the noise suppressing condition for idealistic CDD, we follow similar procedures as the PDD case by applying the error phase approximation Eq. (55),

$$\begin{aligned} \beta_{(n)}^2 &\approx 2^{2n(n+1)} \|\text{ad}_{B_0}^n(B_1)\|^2 \\ &+ 2^{2n^2} \|\text{ad}_{B_0}^n(B_2) + i \text{ad}_{B_0}^{n-1}(\{B_1, B_3\})\|^2 \\ &\leq 4^{n(n+1)} \alpha^{2n-2} (4^n \alpha^2 \|B_1\|^2 + (\alpha \|B_2\| + \|B_1\| \|B_3\|)^2). \end{aligned} \quad (60)$$

Defining  $x_i \equiv \|B_i\|/\beta$ , the noise reduction condition  $\beta_{(n)}/\beta < 1$  can be converted to the optimization problem:

$$\max_{x_1^2 + x_2^2 + x_3^2 = 1} 4^n \alpha^2 x_1^2 + (\alpha x_2 + \beta x_1 x_3)^2 \leq \frac{1}{\alpha^{2n-2} 4^{n(n+1)}}. \quad (61)$$

This inequality is satisfied when:

$$\begin{aligned} \beta/\sqrt{4^n - 1} &\leq \alpha \leq 1/2^{n+2} \\ \text{or } \left( \frac{(4^n - 1)\alpha^2 + \beta^2}{2\alpha\beta} \right)^2 + 1 &\leq \frac{1}{4^{n(n+1)}\alpha^{2n}}. \end{aligned} \quad (62)$$

This reduces to the PDD condition Eq. (58) when  $n = 1$  as one may easily check. Moreover, if the error phase  $\beta \leq \frac{1}{4}\sqrt{1 - 4^{-n}} \approx 1/4$ , a realistic condition for small noise, it suffices to use the simplified cutoff for  $\alpha$ ,

$$\alpha \leq \frac{1}{2^{n+2}}. \quad (63)$$

Meanwhile we can establish the absolute convergence region of the Magnus series according to Eq. (24):

$$\alpha + \beta \leq \frac{1}{4^n}. \quad (64)$$

Now the sufficient condition promising a reduction of the error phase is for the two-component norm  $(\alpha, \beta)$  of the initial Hamiltonian to fall in the overlapping region of Eq. (62) and Eq. (64). In general, the radius of the absolute convergence region gets exponentially smaller as

the concatenation level goes up; and the convergence region completely falls in the region of noise removal for  $n \geq 2$ . This seems to suggest that the absolute convergence condition alone is enough. But numerical tests show that the Magnus series would typically converge well beyond the region set by Eq. (64), the noise reduction condition Eq. (62) is more useful in representing the true noise suppressing region. Following a similar procedure as the PDD case, the noise reduction region is numerically scanned and compared with the predicted boundaries in Figure 3 for level 2 and level 3 CDD.

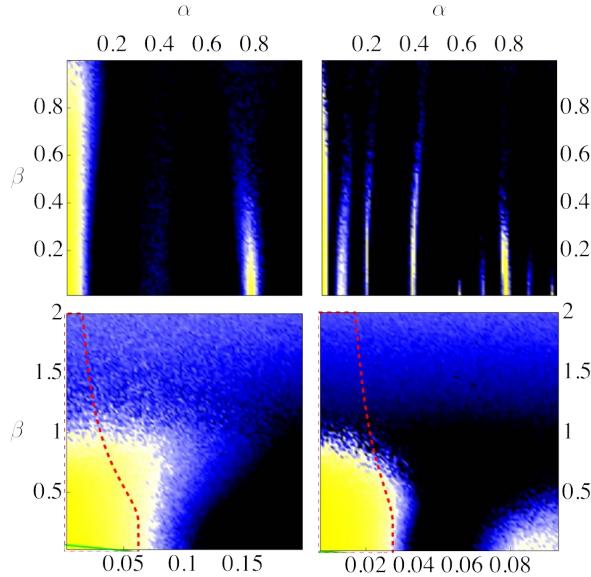


FIG. 3: The noise suppression region (in yellow) for CDD<sub>2</sub> (left) and CDD<sub>3</sub> (right) scanned over the  $(\alpha, \beta)$  parametric plane. Different ranges are used for the upper and lower graphs. The red dotted lines indicated the theoretical suppression boundaries predicted by (62); the green solid line near the origin represents the absolute-convergence condition (64).

From the noise suppressing region in Figure 3, we can conclude that the absolute convergence condition really fails to predict anything useful, especially for higher level CDD where the available region becomes too small—it is an indication of “safeness” but nowhere near the boundary. On the other hand, the bound (62) derived from the leading order Magnus approximation really prevails at reasonably small  $\alpha$  and  $\beta$  values, but fails when  $\beta$  grows larger. In the end, it appears that the simplified bound (63) works the best, as the free evolution noise should be small to start with. Another interesting observation is the yellow spikes at high  $\alpha$  values shown in Figure 3. But such phenomenon is certainly beyond any prediction based on the Magnus series as it lays beyond the radius of converge Predicted by Eq. (64).

## B. Gate errors

Generally, we can denote  $\tilde{P}_i = P_i V_i$  as the noisy implementation of the gate  $P_i$ , with  $V_i \equiv P_i^\dagger \tilde{P}_i$  representing the additional noise associated with the gate. If the noise is unitary, we can associate  $V_i = e^{-i\Gamma_i}$ , where  $\Gamma_i$  is a Hermitian operator on  $\mathcal{H}_S$ . For the CPTP noise case, we can find a unitary embedding by introducing an ancillary system  $\mathcal{H}_A$  (subscript “A” for ancilla), such that  $V_i = e^{-i\Gamma_i}$  for some Hermitian operator  $\Gamma_i$  on the extended Hilbert space  $\mathcal{H}_S \otimes \mathcal{H}_A$ . These two scenarios can be unified if we allow  $\dim(\mathcal{H}_A) = 1$ . The time evolution operator for a noisy DD cycle now becomes:

$$\begin{aligned} \tilde{U}_{\text{DD}} \equiv e^{-i\tilde{\Omega}_{\text{DD}}} &= \tilde{P}_K e^{-i\Omega} \cdots \tilde{P}_2 e^{-i\Omega} \tilde{P}_1 e^{-i\Omega} \\ &= P_K e^{-i\Gamma_K} e^{-i\Omega} \cdots P_2 e^{-i\Gamma_2} e^{-i\Omega} P_1 e^{-i\Gamma_1} e^{-i\Omega}, \end{aligned} \quad (65)$$

where we have associated an effective Hamiltonian  $\tilde{\Omega}_{\text{DD}}$  for the noisy DD cycle. For simplicity we make the assumption that  $\mathcal{H}_A$  and  $\mathcal{H}_B$  are disjoint Hilbert spaces. Then  $\tilde{\Omega}_{\text{DD}}$  can be recognized as a Hermitian operator on the combined space  $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_B$ . We also denotes its two-component norm by  $\|\tilde{\Omega}_{\text{DD}}\| = (\tilde{\alpha}_{\text{DD}}, \tilde{\beta}_{\text{DD}})$ . This can be done by viewing  $\mathcal{H}_A \otimes \mathcal{H}_B$  as the new bath Hilbert space. 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 now establish the noise threshold criteria for noisy DD gates by solving the error phase inequality  $\tilde{\beta}_{\text{DD}} \leq \beta$ . In the followings, We split our discussions according to the noise model considered.

For gate-independent noise,  $\tilde{P}_i = P_i V = P_i e^{-i\Gamma}, \forall i$ , a great simplification can be achieved in Eq. (65) by combining the gate-noise with the free-evolution noise into an effective free evolution map:

$$\tilde{U}_f = e^{-i\Gamma} e^{-i\Omega} \equiv e^{-i\tilde{\Omega}}, \quad (66)$$

where we have associate an effective Hamiltonian  $\tilde{\Omega}$  for the effective free evolution  $\tilde{U}_f$ . To calculate the effective Hamiltonian  $\tilde{\Omega}_{\text{DD}}$  after such noisy DD, we can simply replace the free evolution Hamiltonian  $\Omega$  with  $\tilde{\Omega}$  and apply the same formula for calculating the idealistic DD.

It is thus imperative to understand how the gate error enters  $\tilde{\Omega}$ . Analytic expression for  $\tilde{\Omega}$  does not exist for non-commuting matrices. We can only express as it an infinite series by invoking the BCH formula,

$$\tilde{\Omega} = \Gamma + \Omega - \frac{i}{2} [\Gamma, \Omega] - \frac{1}{12} ([\Gamma, [\Gamma, \Omega]] - [\Omega, [\Gamma, \Omega]]) + \cdots,$$

where higher order terms are nested commutators in  $\Gamma$  and  $\Omega$ . Just as  $\Omega$  can be decomposed in to  $\Omega_B + \Omega_{\text{SB}}$ , we can apply a similar decomposition for  $\Gamma$  as  $\Gamma_A + \Gamma_{\text{SA}}$ . To characterize the gate error, we define its two-component norm by

$$\|\Gamma_A\| = \eta_0, \quad \|\Gamma_{\text{SA}}\| = \eta. \quad (67)$$

Per assumption, the ancilla and bath Hilbert spaces are disjoint. Therefore,  $[\Omega_B, \Gamma] = [\Gamma_A, \Omega] = 0$ . This indicates that both  $\Omega_B$  and  $\Gamma_A$  would vanish in the BCH series except for in the linear term, where they contribute to the pure bath part of  $\tilde{\Omega}$ . Defining  $W \equiv \tilde{\Omega} - \Gamma_A - \Omega_B$ , it then follows that  $\mathcal{P}_c(W) = \mathcal{P}_c(\tilde{\Omega})$  and  $e^{-iW} = e^{-i\Gamma_{SA}}e^{-i\Omega_{SB}}$ . As a result, we can bound the new error phase by

$$\begin{aligned}\tilde{\beta} &\equiv \|\mathcal{P}_c(\tilde{\Omega})\| = \|\mathcal{P}_c(W)\| \leq \|W\| \\ &= \|U_1\Gamma_{SA}U_1^\dagger + U_2\Omega_{SB}U_2^\dagger\| \leq \|\Gamma_{SA}\| + \|\Omega_{SB}\| \\ &= \eta + \beta,\end{aligned}\quad (68)$$

where we have applied the norm contraction property of the projection map  $\mathcal{P}_c$  on to the coupling space; the Thompson's formula [25] in line 2, which promises the existence of unitary matrices  $U_1$  and  $U_2$  such that  $W = U_1\Gamma_{SA}U_1^\dagger + U_2\Omega_{SB}U_2^\dagger$ . On the other hand, as commutator of the traceless Pauli matrices remain traceless,  $[\Gamma_{SA}, \Omega_{SB}]$  has no pure bath contribution. By expanding the series up to the third order, we can show

$$\tilde{\alpha} \leq \alpha + \eta_0 + \frac{1}{3}\eta\beta^2 + \mathcal{O}((\eta, \beta)^4). \quad (69)$$

For small noise, it suffice to use the third order correction as an upper bound for  $\tilde{\alpha}$ .

These two bounds on  $\tilde{\alpha}$  and  $\tilde{\beta}$  can be used to derive the generic noise suppressing condition by the method of simple substitution. Specifically, let's say that we have derived the error phase upper bound for idealist DD as a polynomial  $f(\alpha, \beta)$  such that  $\beta_{DD} \leq f(\alpha, \beta)$  for all free evolution Hamiltonian specified by  $(\alpha, \beta)$ . This inequality would still apply for noisy DD given effective Hamiltonian specified by  $(\tilde{\alpha}, \tilde{\beta})$ . To compare with the free evolution case, we then upper-bound the maximal error phase after noisy DD by  $f(\tilde{\alpha}, \tilde{\beta}) \leq \beta$ , now with the substitution  $\tilde{\alpha} = \alpha + \eta_0 + \frac{1}{3}$  and  $\tilde{\beta} = \beta + \eta$  as shown in Eq. (68) and Eq. (69). This generic method of deriving threshold is conceptually simple and applies all DD schemes, but the resulting bound could be suboptimal. Particularly when there is nested commutators in the expression of the DD estimator. To derive a better threshold, detailed analysis should be performed with respect to the specific DD scheme.

### C. Gate independent noise

Let us consider the noise suppressing condition for CDD under the gate-independent unitary noise model, where an over-rotation error is associated with all gates,  $\tilde{P}_i = P_i e^{-i\vec{\sigma} \cdot \vec{\theta}}$ , where  $\vec{\theta} = (\theta_1, \theta_2, \theta_3)^\top$ . There is no pure bath part here as it only contributes to a global phase, the error phase is given by the rotation angle  $\|\vec{\theta}\| \equiv \theta$ . Such type of noise could happen, for example, if their is a systematic calibration error in the system.

Let us expand the write the effective free evolution Hamiltonian from Eq. (66) as:

$$\tilde{\Omega} = \vec{\sigma} \cdot \vec{\theta} + \Omega' = \sigma_0 \otimes B'_0 + \sum_{i=1}^3 \sigma_i \otimes (\theta_i + B'_i). \quad (70)$$

Up to second order in the BCH series, we have  $B'_0 \approx B_0$  and  $B'_i \approx B_i + \sum_{jk} \epsilon_{ijk} \theta_j B_k$ ,  $i = 1, 2, 3$ . Up to first order smallness in  $\theta$ , the expression for  $(B'_1, B'_2, B'_3)$  can obtained by rotating the vector  $(B_1, B_2, B_3)$  with respect to the  $\vec{\theta}$ -axis. 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  $\Omega'$  remains identical to  $\beta$  up to the second order in  $\theta$ . To conclude, the two-component norm of  $\Omega'$  can be estimate by

$$\|\Omega'\| = \left( \frac{\tilde{\alpha}}{\beta'} \right) \leq \left( \frac{\alpha + \frac{1}{3}\theta\beta^2}{\beta + \theta^2\beta} \right). \quad (71)$$

Assuming the noise to be small, we may only use the leading order approximation, knowing that it is accurate to the second order.

We now calculate the noise suppression condition for  $CDD_n$ , with PDD recognized as the special case of  $n = 1$ . From the formula for the  $CDD_n$  estimator in Eq. (43) and the free evolution Hamiltonian  $\tilde{\Omega}$  in Eq. (70), we can estimate the coupling part of  $\tilde{\Omega}_{(n)}$  as

$$\begin{aligned}\mathcal{P}_c(\Omega_{(n)}) &\approx \sigma_1 \otimes (-i)^n 2^{n(n+1)} \text{ad}_{B'_0}^n(B'_1) \\ &+ \sigma_2 \otimes (-i)^n 2^{n^2} \text{ad}_{B'_0}^{n-1}([B'_0, B'_2] - i\{B'_1 + \theta_1, B'_3 + \theta_3\}),\end{aligned}\quad (72)$$

where all numbers  $\theta_i$  in the commutators have been eliminated. We proceed to estimate the error phase upper-bound using the approximation formula Eq. (55) and applying triangular inequality for operator norm:

$$\begin{aligned}\beta_{(n)}^2 &\leq 2^{2n(n+1)} \|B'_0\|^{2n-2} \left\{ (2^n \|B'_0\| \|B'_1\|)^2 \right. \\ &\quad \left. + (\|B'_0\| \|B'_2\| + (\theta_1 + \|B'_1\|)(\theta_3 + \|B'_3\|))^2 \right\}.\end{aligned}\quad (73)$$

Upper-bounding the r.h.s. by  $\beta^2$ , we eventually arrive at the optimization problem:

$$\begin{aligned}\max_{\substack{x_i, y_i \\ \|\vec{x}\| = \|\vec{y}\| = 1}} (2^n \tilde{\alpha} x_1)^2 + \left( \tilde{\alpha} x_2 + \frac{(\theta y_1 + \beta x_1)(\theta y_3 + \beta x_3)}{\beta} \right)^2 \\ \leq \frac{1}{2^{2n(n+1)} \tilde{\alpha}^{2n-2}},\end{aligned}\quad (74)$$

where we have recognized  $\|B'_0\| \leq \alpha + \beta^2\theta/3 \equiv \tilde{\alpha}$  and defined  $x_i \equiv \|B'_i\|/\beta$ ,  $y_i \equiv \theta_i/\theta$ . The normalization constraint for  $\vec{x}$  is set to  $\|\vec{x}\| = 1$  in stead of the original  $\|\vec{x}\| = \beta'/\beta$ . According to Eq. (71), the deviation for this smaller norm is on the order of  $\mathcal{O}(\theta^2)$  which can be safely ignored as higher order smallness.

The analytical solution to this problem is not known to us, but we can instead solve a slightly weaker condition given all parameters are reasonably small. Next we bound the term involving  $\vec{y}$  by

$$(\theta y_1 + \beta x_1)(\theta y_3 + \beta x_3) \leq \frac{1}{2}\theta^2 + \sqrt{x_1^2 + x_3^2}\theta\beta + \beta^2x_1x_3,$$

which can be shown by expanding the bracket and maximizing the coefficients individually. We also realize that the maximization should be achieved at around  $x_2 = 0$  to take advantage of either a large  $\tilde{\alpha}$  or a large  $\beta$ . This leads to the slightly weaker maximization problem:

$$\max_{x_1^2+x_3^2=1} (2^n\tilde{\alpha}x_1)^2 + \left(\frac{\theta^2}{2\beta} + \theta + \beta x_1 x_3\right)^2 \leq \frac{1}{4^{n(n+1)}\tilde{\alpha}^{2n-2}}, \quad (75)$$

whose solution can be now be found analytically, albeit complicated. Now given a specific  $\theta$  value, this inequality would determine a continuous region in the  $(\alpha, \beta)$ -plane where noise suppression is possible. On the other hand, for a particular pair of  $(\alpha, \beta)$  values, this inequality would determine a maximally allowed  $\theta$ .

Let us first focus on PDD. In Figure 4, We illustrate the noise suppression regions obtained from (75) for PDD with increasing levels of noise strength  $\theta$ . This figure also intuitively demonstrates how the maximally allowed  $\theta_{\max}(\alpha, \beta)$  depends on the parameters  $(\alpha, \beta)$ . In general, a smaller  $\alpha$  value leads to a larger  $\theta_{\max}$ . An optimal spot in the phase plane which is able to tolerate most error can be found at around near the  $\alpha$ -axis.

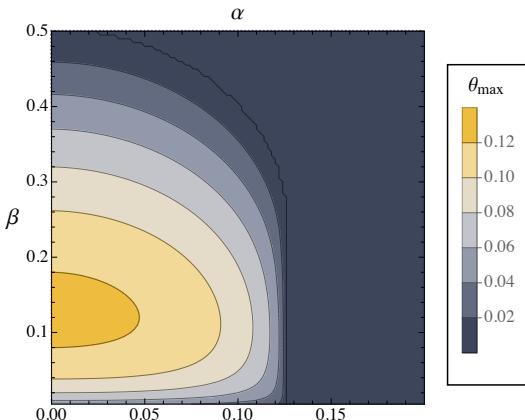


FIG. 4: The noise suppressing region for noisy PDD with varying levels of noise  $\theta$ . Also can be understood as the contour plot of the maximally tolerable noise level  $\theta_{\max}$ .

Another observation from Figure 4 is that the lower-left corner around the origin gradually leaves the noise suppressing region as the gates get noisier. This is because the noise associated with free evolution is small compared with the noise in the gates ( $\beta \ll \theta$ ), so that

the cost of noisy DD overwhelms its benefit. Thus this regime of is important for the purpose of deriving the gate error threshold. For  $\beta \ll \theta$ , the second parenthesis in Eq. (75) will dominate due to the small denominator. With the same logic, this approximation also applies to the regime where  $\tilde{\alpha} \ll \beta$ . In those cases, the optimization can then be approximately achieved by only considering maximizing the second term. This is achieved when  $x_1 = x_3 = 1/\sqrt{2}$  independent of the values of  $\theta$  and  $\beta$ . Substituting this solution back to Eq. (75), then we can predict the following simplified bound for PDD:

$$2\tilde{\alpha}^2 + \frac{(\theta + \beta)^4}{4\beta^2} \leq \frac{1}{16}, \quad (76)$$

From this relation we can solve for  $\theta$  in terms of  $\alpha$  and  $\beta$ . After keeping the leading order terms, we arrive at the noise threshold for the rotation angle:

$$\theta \leq (1 - 32\alpha^2)^{\frac{1}{4}} \sqrt{\frac{\beta}{2}} - \beta \lesssim \sqrt{\frac{\beta}{2}}. \quad (77)$$

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 for PDD to be considered worthwhile. Further more, from this inequality, it is clear that the maximally allowed  $\theta$  is a monotonically decreasing function in  $\alpha$ . The globally maximally allowed  $\theta$  can be found at  $\alpha = 0$ ,  $\beta = 1/8$  with  $\theta_{\max} = 1/8$ .

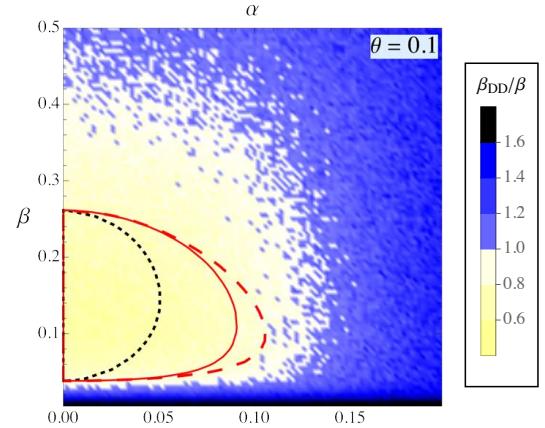


FIG. 5: The maximal  $\beta_{\text{DD}}/\beta$  ratio obtained from numerical simulations for noisy PDD with  $\theta = 0.1$ . Boundaries of the theoretic noise suppressing regions are marked for comparison: red solid line indicates result from optimization Eq. (74), black dotted line stands for the generic boundary after simple substitution, and the red dashed line represents the approximate condition Eq. (76).

To examine the validity of our results, we numerically simulate PDD with fixed unitary error at  $\theta = 0.1$ . The simulation is performed in the similar manner as the ideal

PDD case except with the additional errors. The data is presented in Figure 5, together with the theoretical boundary predicted by Eq. (75) marked in red solid line. For comparison, the noise suppression bound from simple substitute of the ideal PDD condition with  $\alpha \rightarrow \tilde{\alpha}$  and  $\beta \rightarrow \tilde{\beta}$  is plotted in dashed black line; and the simplified bound Eq. (76) is also plotted as dashed red line. We see that the three boundaries coincides at  $\alpha = 0$  while extending differently into the parameter plane. The simple substitution strategy produces the smallest region. The optimized boundary from Eq. (75) gives a better result after taking into account of the commutator terms. The approximate condition Eq. (76) works well in capturing the  $\alpha \ll \theta$  behavior, but is already violated by numerical evidences from Figure 5 for larger  $\alpha$ , as indicated by the blue dots within its boundary. Another observation from Figure 5 is the relative insensitivity of the  $\alpha$  value with respect to the noise suppression region, as indicated by the bottom boundary which is almost parallel to the horizontal axis. This insensitivity is also reflected in the error threshold condition Eq. (77).

As is the case for PDD, the trial solution  $x_1 = x_3 = 1/\sqrt{2}$  for Eq. (75) still gives a decent approximate solution for higher level CDD. After substitution, we have:

$$2^{2n-1}\tilde{\alpha}^2 + \frac{(\theta + \beta)^4}{4\beta^2} \leq \frac{1}{4^{n(n+1)}\tilde{\alpha}^{2n-2}}, \quad (78)$$

This inequality establishes the region where the error phase is reduced by CDD according to the leading order Magnus expansion. In Figure 6, we plot the noise removal region determined by Eq. (78) for CDD<sub>2</sub> and CDD<sub>3</sub> with different levels of noise  $\theta$ . Instead of plotting the entire phase space, we focus on the regime where  $\alpha$  and  $\beta$  are small, as defined by Eq. (63), so that our approximate condition is accurate.

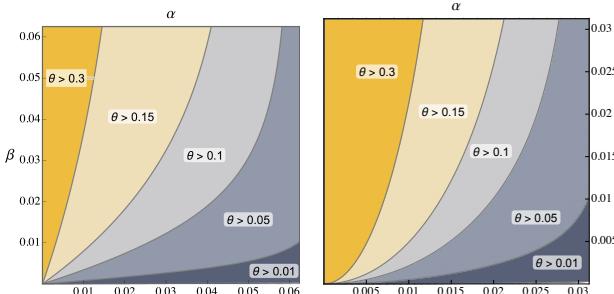


FIG. 6: The noise removal region of CDD<sub>2</sub> (left) and CDD<sub>3</sub> (right) for different levels of noise strength specified by  $\theta$ . Notice the ranges for  $\alpha$  and  $\beta$  are different for different concatenation levels.

Some differences can be seen by comparing Figure 4 and Figure 6. In the PDD case, the noise reduction region quickly shrinks to a point as  $\theta$  increases. But this is not the case for higher level CDD, where the region is pushed closer and closer to the  $\alpha = 0$  boundary. Also as opposed to the PDD, there is no globally maximal  $\theta_{\max}$

predicted from the noise suppressing condition Eq. (78). This difference is most obvious by considering the  $\alpha = 0$  limit. According to Eq. (78), the r.h.s. blows up for  $n \geq 2$ , implying very large noise can be tolerated. The counter-intuitive prediction of noise tolerance is clearly incorrect. The fault stems from failure in approximating the Magnus series by its leading order term. In Eq. (43) for the leading order term of the CDD effective Hamiltonian, all coupling part bath operators are expressed as nested commutators with  $B_0$  for  $n \geq 2$ . A vanishing  $B_0$  therefore implies the diminish of the entire coupling part, leaving the next order term to be “leading”. This argument also applies when  $B_0 \neq 0$ : while the error phase of the leading order term is  $\mathcal{O}(\alpha^n\beta)$ , the next order term contains component on the order of  $\beta^{n+2}$ . Now that if  $\alpha \ll \beta$ , there is no way to be sure that  $\alpha^n\beta \gg \beta^{n+2}$ , which suggests that the higher order term could turn out to be larger than the estimator. Therefore, to correctly estimate the error phase of CDD and derive the noise threshold, we need to assume that  $\alpha \neq 0$  and that  $\alpha$  and  $\beta$  are on the same order of smallness.

We now assume that  $\alpha$  and  $\beta$  are of similar size. Then the approximation  $\tilde{\alpha} \approx \alpha$  can be made. From Eq. (78), we can solve for  $\theta$  and keep the leading order terms in  $\beta$ . The following threshold can be obtained:

$$\theta \leq \left( (2^{n+2}\alpha)^{2(1-n)} - \frac{1}{2}(2^{n+2}\alpha)^2 \right)^{\frac{1}{4}} \sqrt{\frac{\beta}{2}} - \beta. \quad (79)$$

Supposing that ideal CDD<sub>n</sub> is allowed, we can recognize the cutoff  $(2^{n+2}\alpha) \leq 1$  from Eq. (63). Since the coefficient before  $\sqrt{\beta/2}$  is a monotonically decreasing function in  $(2^{n+2}\alpha)$ , which attains the minimal value at the cutoff. We can express the threshold in terms of the error phase alone:

$$\theta \leq (1/8)^{\frac{1}{4}} \sqrt{\beta} - \beta. \quad (80)$$

This threshold shows that for CDD the gate noise level should still be bounded by the square root of the free evolution noise level. This is consistent with our finding for the PDD case. One the other hand, for fixed free evolution Hamiltonian (therefore fixed  $\alpha, \beta$ ), different CDD concatenation levels could turn out to have different noise thresholds due to the leading coefficient in Eq. (79) as a function of  $n$ . It turns out that there exists an optimal concatenation level in terms of the noise threshold. Let us say that  $\alpha \approx 1/2^{N+2}$ ,  $N \geq 2$ , so that the maximal allowed concatenation level is  $N$ . Substituting back to Eq. (79) and focusing on the term within the parenthesis, the following approximation can be made:

$$2^{2(n-N)(1-n)} - \frac{1}{2}2^{2(n-N)} \approx 2^{2(N-n)(n-1)}, \quad 1 \leq n \leq N.$$

After maximizing the exponent, we find the optimal concatenation level is at  $(N+1)/2$ . This calculation suggests that increasing the concatenation level to the maximal allowed level may not be a good idea with noisy gates. Knowing the  $\alpha$  value, it is possible to perform optimization on the coefficient to obtain the best concatenation

level. Otherwise, one is advised to follow the generic threshold Eq. (80).

#### D. gate-dependent noise

Let us consider the case where noise is dependent on the gate.

For PDD and CDD, as the sequence is composed of only the elementary  $X$  and  $Z$  gates, we assume that the noisy implementation of these gates are

$$\tilde{X} = X e^{-i\Gamma'}, \quad \tilde{Z} = Z e^{-i\Gamma''}. \quad (81)$$

With these noisy gates substituted, the PDD time evolution map can be written as

$$\begin{aligned} e^{-i\tilde{\Omega}_P} &= Ze^{-i\Omega''} X e^{-i\Omega'} Ze^{-i\Omega''} X e^{-i\Omega'} \\ &= e^{-iZ\Omega''} Z e^{-iY\Omega' Y} e^{-iX\Omega'' X} e^{-i\Omega'}, \end{aligned} \quad (82)$$

where we have defined

$$e^{-i\Omega'} \equiv e^{-i\Gamma'} e^{-i\Omega}, \quad e^{-i\Omega''} \equiv e^{-i\Gamma''} e^{-i\Omega}. \quad (83)$$

We can now apply the Magnus formula Eq. (25) to solve for the PDD effective Hamiltonian  $\tilde{\Omega}_P$ . Supposing that  $\Omega' = \sum_i \sigma_i \otimes B'_i$  and  $\Omega'' = \sum_i \sigma_i \otimes B''_i$ , the first two Magnus terms are now

$$\tilde{\Omega}_P^{(1)} = \sigma_0 \otimes 2(B'_0 + B''_0) + \sigma_2 \otimes 2(B'_2 - B''_2) \quad (84)$$

$$\begin{aligned} \tilde{\Omega}_P^{(2)} &= \sigma_0 \otimes i([B'_0, B''_0] + [B'_1, B''_1] - [B'_2, B''_2] - [B'_3, B''_3]) \\ &+ \sigma_1 \otimes (-i[B'_0 + B''_0, B'_1 + B''_1] + \{B'_2 - B''_2, B'_3 - B''_3\}) \\ &+ \sigma_2 \otimes (-i[B'_0, B''_2] - i[B''_0, B'_2] - \{B'_1, B''_3\} - \{B''_1, B'_3\}) \\ &+ \sigma_3 \otimes (i[B'_0 + B''_0, B''_3 - B'_3] + \{B'_1 + B''_1, B''_2 - B'_2\}). \end{aligned} \quad (85)$$

When  $\Omega' = \Omega''$ , these two equation would reduce to the expressions for the idealist PDD, which achieve first order decoupling. However with gate-dependent noise, the coupling part of the first order Magnus term no longer vanishes. In case where the gate noise strength is at least on the same level of the free evolution noise, the error phase of the PDD effective Hamiltonian is dominated by the difference of  $\Omega'$  and  $\Omega''$ . From Eq. (83) and the BCH series, we may estimate:  $\Omega' - \Omega'' \approx (\Gamma_1 - \Gamma_2) - \frac{i}{2}[\Gamma_1 - \Gamma_2, \Omega]$ . Since the noise for  $X$  and  $Z$  gate could be totally independent, the difference of  $\Omega'$  and  $\Omega''$  is dominated by the first order term:

$$\mathcal{P}_c(\tilde{\Omega}_P^{(1)}) \simeq \sigma_2 \otimes 2(A'_2 - A''_2). \quad (86)$$

So the error phase upper bound of the PDD effective Hamiltonian is estimated to be

$$\tilde{\beta}_P \approx \|2(A'_2 - A''_2)\| \leq 4\eta, \quad (87)$$

where  $\eta$  is the maximal error phase of the gate noise:

$$\eta \equiv \max(\|\mathcal{P}_c(\Gamma_1)\|, \|\mathcal{P}_c(\Gamma_2)\|). \quad (88)$$

The noise threshold, according to our definition, is the maximally allowed  $\eta$  such that the error phase is still reduced through DD. Therefore for our model, it is simply given by

$$\eta \leq \beta/4. \quad (89)$$

Let us consider the CDD case if the higher level map is constructed from the noisy lower level maps following the recursive definition Eq. (36). Same as the idealistic case, for  $CDD_n$ , the effective Hamiltonian is produced by repeated application of the PDD map:

$$\tilde{\Omega}_{(n)} = \underbrace{\tilde{\Omega}_P \circ \dots \circ \tilde{\Omega}_P}_{n \text{ folds}}(\Omega), \quad (90)$$

where  $\tilde{\Omega}_P$  is the map on  $\Omega$  of the noisy PDD, with the first two orders now given by Eq. (84) and Eq. (85). But unlike the idealist or gate-independent noise case, the first order decoupling is no longer possible. Therefore the decoupling order of  $CDD_n$  would break down as well, with the first order term persisting despite the concatenation level. To estimate the first order term, it suffice to use the estimator  $\tilde{\Omega}_P^{(1)} \circ \dots \circ \tilde{\Omega}_P^{(1)}(\Omega)$ . According to Eq. (86), the coupling part of the PDD map is independent of  $\Omega$  in the leading order, therefore  $\mathcal{P}_c(\tilde{\Omega}_{(n)}^{(1)}) \simeq \sigma_2 \otimes 2(A'_2 - A''_2)$  as well. The error phase after  $CDD_n$  is still approximated by

$$\tilde{\beta}_{(n)} \approx \|2(A'_2 - A''_2)\| \leq 4\eta, \quad \forall n. \quad (91)$$

Therefore the same error threshold Eq. (89) would still apply for CDD independent of the concatenation level.

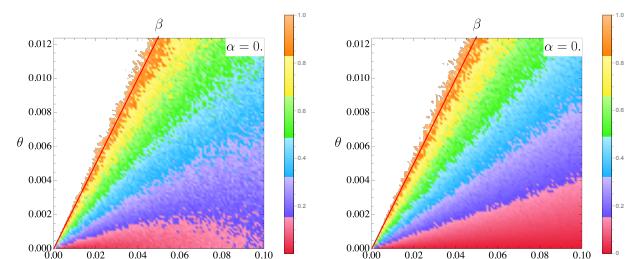


FIG. 7: The noise threshold compared for PDD and  $CDD_3$  simulated at  $\alpha = 0$ . The red solid line represents the noise threshold prediction  $\eta \leq \beta/4$ . Different colors represents different levels of noise reduction ratio  $\beta_{DD}/\beta$  after numerically maximized over random bath operator instances.

In Figure 7, we simulate both PDD and  $CDD_3$  with the gate-dependent noise model. We find that our prediction works pretty well.

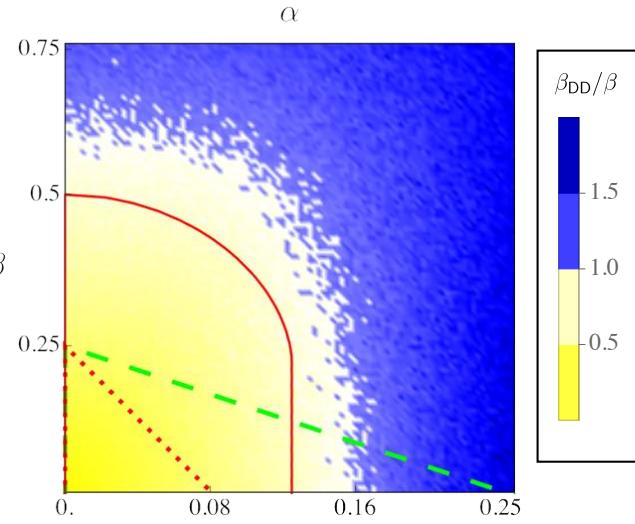
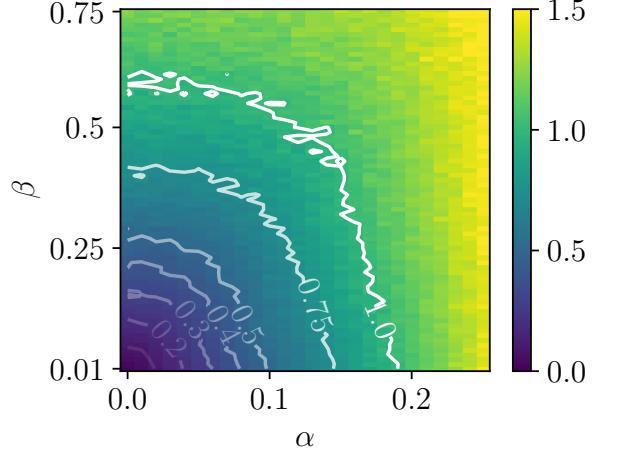


FIG. 8: Reproduction of Fig. 2 by using error phase with operator norm

## V. IMPACT OF DIFFERENT MEASURES

### A. Error phase benchmarking

We reproduce the error phase results in Fig. 2 in this subsection to make sure that the generation of the Hamiltonian and the usage of the norm are consistent. In Fig. 8(a), The region has similar boundaries as compared In Fig. 8(b) (Fig. 2)

### B. Error phase with different norm usage

Apart from the error phase with operator norm, we also use the Frobenius norm (Hilbert-Schmidt norm) to compute the error phase though this is inconsistent with the previous definition. This is to setup possible reference with respect to the channel infidelity measure as  $\text{InF}(\mathcal{E}) \simeq$

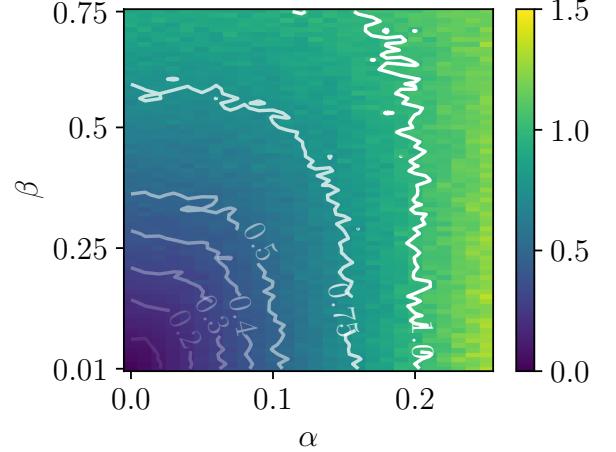


FIG. 9: Error phase with Hilbert-Schmidt norm

$$\|\Omega_{\text{SB}}\|_\infty^2.$$

### C. Channel infidelity and trace distance

We proceed to consider to check channel infidelity as a possible measure whether PDD would work as expected. The detailed implementation is given in Appendix D. Note that the definition is given by for a specify channel  $\mathcal{E}$ ,

$$D_{\max}(\mathcal{E}) = \max_{\rho \in \mathcal{D}(\mathcal{H})} \frac{1}{2} \|(\mathcal{I} - \mathcal{E})(\rho)\|_1 \equiv \frac{1}{2} \|\mathcal{I} - \mathcal{E}\|_1, \quad (92)$$

$$\text{InF}_{\max}(\mathcal{E}) \equiv \max_{\rho \in \mathcal{D}(\mathcal{H})} \text{tr}(\rho (\mathcal{I} - \mathcal{E})(\rho)). \quad (93)$$

However, for a set of  $\alpha$  and  $\beta$ , there could be various different channels  $\mathcal{E}_i$ . We have therefore take the maximum over these channels as well thus

$$D_{\max} = \max_{\mathcal{E}_i, \rho \in \mathcal{D}(\mathcal{H})} \frac{1}{2} \|(\mathcal{I} - \mathcal{E}_i)(\rho)\|_1 \equiv \frac{1}{2} \|\mathcal{I} - \mathcal{E}\|_1, \quad (94)$$

$$\text{InF}_{\max} \equiv \max_{\mathcal{E}_i, \rho \in \mathcal{D}(\mathcal{H})} \text{tr}(\rho (\mathcal{I} - \mathcal{E}_i)(\rho)). \quad (95)$$

We restrict ourselves to the infidelity of PDD channel and free evolution channel, namely  $\text{InF}_{\max}^{\text{DD}}(\beta, \alpha, \eta)$ ,  $\text{InF}_{\max}^{\text{free}}(\beta, \alpha)$ .

We thus consider the ratio of the channel infidelity for DD sequence and that for free evolution, i.e.,  $\text{InF}_{\max}^{\text{DD}} / \text{InF}_{\max}^{\text{free}}$ .

### D. Difference measures versus $\beta$

These measures show very different error suppression boundaries. We consider a give measure  $\mathcal{M}$  for DD and free evolution respectively and inspect how it scale with  $\beta$ . From Fig. 12, the error for free evolution (solid lines)

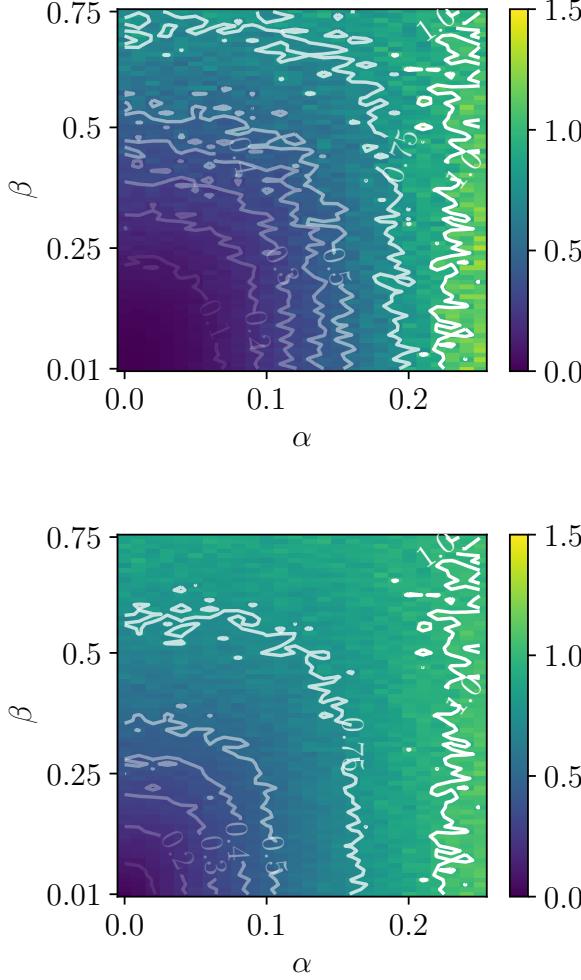


FIG. 10: Ratio of  $\ln F_{\max}^{\text{DD}} / \ln F_{\max}^{\text{free}}$  and its square root. We consider the ideal situation where the pulses are ideal. For each pair of  $\alpha$  and  $\beta$ , we consider 10000 samples channels  $\mathcal{E}_i$

are all linear with respect to  $\beta$ , with different gradients. Channel infidelity for free evolution (blue solid line) is quadratic with respect to  $\beta$  which has been proven in Appendix A. Channel trace distance (orange solid lines), on the other hand, is linearly related to  $\beta$ . This suggests a different different error tolerance for these measures. For example, free evolution channel trace distance has reached the upper bound defined by Fuchs-van de Graaf inequalities and thus is bounded by the sqrt root of infidelity showing a linear dependence with respect to  $\beta$ . This would imply a large error for free evolution. On the other hand, the channel trace norm with PDD is quadratic with respect to  $\beta$  corresponding to the free evolution infidelity. (Qns: To what extent do they correspond to just a scaling on the parameters? ). For error phase using operator norm and Frobenius norm, both of them are linear with respect to  $\beta$  for free evolutions.

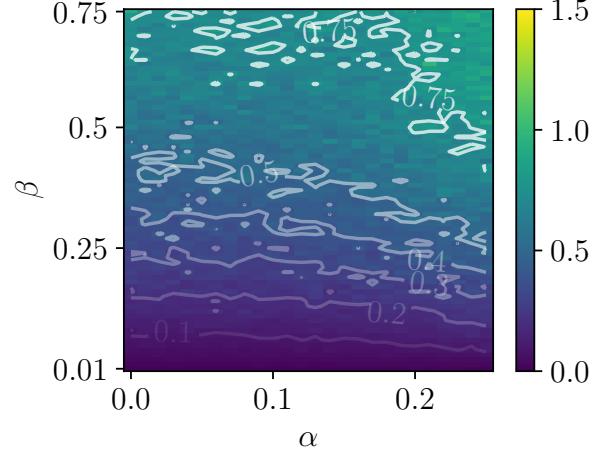


FIG. 11: Ratio of  $D_{\max}^{\text{DD}} / D_{\max}^{\text{free}}$ . We consider the ideal situation where the pulses are ideal. For each pair of  $\alpha$  and  $\beta$ , we consider 10000 samples channels  $\mathcal{E}_i$

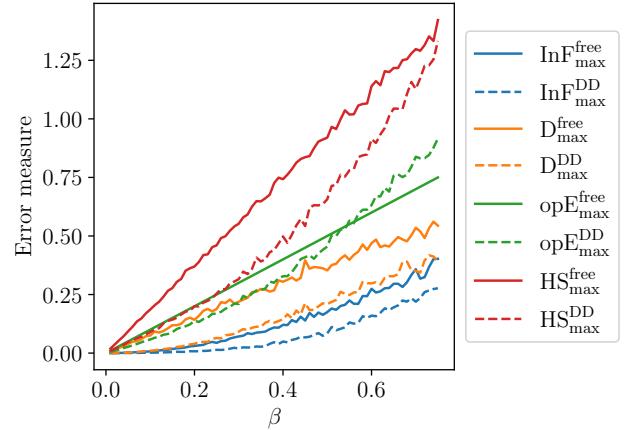


FIG. 12: Different error measures versus  $\beta$  at a fix  $\alpha$ .  $\alpha = 0.1$ . Free evolution with  $\Omega$  (solid lines) and evolution under PDD cycle  $\Omega_P$  (dashed lines) by using channel infidelity (blue), channel trace distance (orange), operator norm (green) (error phase), and Hilbert-Schmidt norm (red).

## E. Noisy pulse

By using different measures, we do see a different noise reduction region. This is demonstrated by the comparison between noise suppression region using error phase and channel infidelity in Fig. 13. The noise suppression region by channel infidelity is larger than that of the error phase.

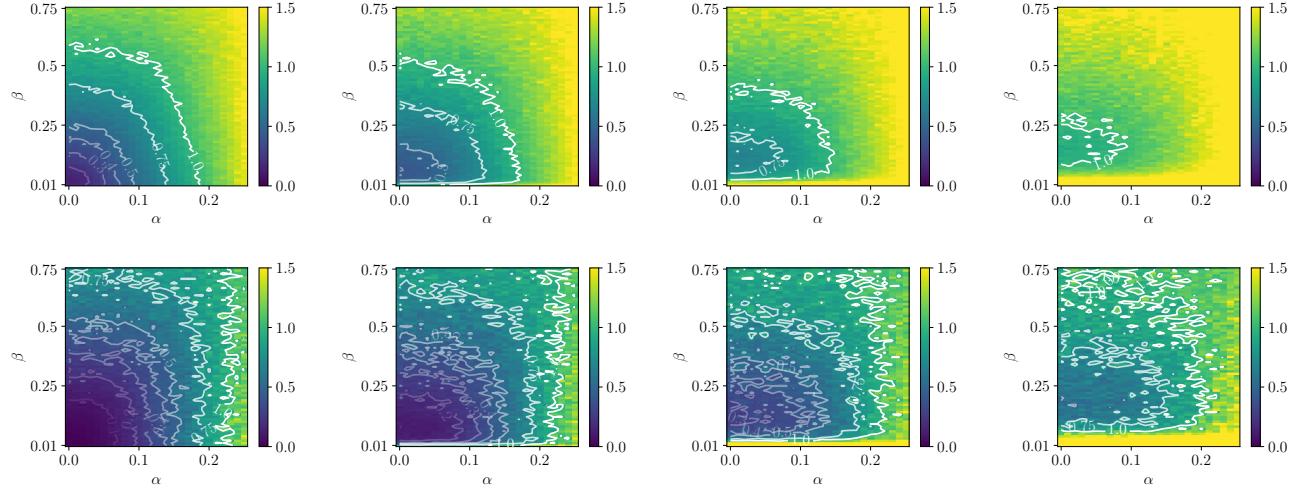


FIG. 13: Top panels: reproduction of the error phase results in Fig. 4. Bottom panels: Noise reduction region using channel infidelity. The strength of noise in the pulse  $\theta = 0.02, 0.06, 0.1, 0.14$  from left to right.

## VI. CONCLUSION

The conclusion that CDD is superior to PDD by just looking at Figure ?? should be taken with a grain of salt. In general, for CDD, despite the fact that the noise decoupling order increases with the concatenation level, it might not prove helpful for realistic noisy removal purposes. From Figure ??, we find that the absolute-convergence region rapidly shrink with the concatenation level. This means that for a system whose  $(\alpha, \beta)$  are not *really* small to start with, our analysis may not

work anymore with high concatenation levels due to convergence issue. If we consider instead the overlapping region between Figure ?? and Figure ??, the noise removal criteria as determined by Eq. (64) and Eq. (62) become increasingly stringent as the concatenation level increases. In general, high-order CDD may not be a good noise suppression choice for quantum computation. To determine the suitable concatenation level, one needs to weigh the costs and benefits between introducing more errors by more pulses and the better noise removal capability promised by of higher order CDD .

- 
- [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 \[quant-ph\]](https://arxiv.org/abs/quant-ph/9807072).
- [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] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Physical Review Letters **77**, 198 (1996).
- [13] D. Gottesman, arXiv preprint quant-ph/9705052 (1997).
- [14] E. Knill and R. Laflamme, Physical Review A **55**, 900 (1997).
- [15] E. Knill, R. Laflamme, and L. Viola, Physical Review Letters **84**, 2525 (2000).
- [16] P. W. Shor, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, 1996) pp. 56–65.
- [17] D. Gottesman, Physical Review A **57**, 127 (1998).
- [18] D. Aharonov and M. Ben-Or, SIAM Journal on Computing (2008).
- [19] J. R. West, D. A. Lidar, B. H. Fong, and M. F. Gyure, Phys. Rev. Lett. **105**, 230503 (2010).
- [20] W. Rossmann, *Lie groups: an introduction through linear groups*, Vol. 5 (Oxford University Press on Demand, 2006).
- [21] E. L. Hahn, Phys. Rev. **80**, 580 (1950).
- [22] H. Y. Carr and E. M. Purcell, Physical review **94**, 630 (1954).
- [23] S. Meiboom and D. Gill, Review of scientific instruments **29**, 688 (1958).
- [24] S. Blanes, F. Casas, J. Oteo, and J. Ros, Physics reports **470**, 151 (2009).

- [25] R. C. Thompson, Linear and Multilinear Algebra **19**, 187 (1986).

### Appendix A: Channel infidelity series in error phase

In the following, we show that the channel infidelity function Eq. (D8) for a CPTP channel does not contain terms linear in the error phase when written as a infinite series. Alternative, by writing  $\Omega = \alpha\Omega_B + \beta\Omega_{SB}$  for real numbers  $\alpha$  and  $\beta$ , we need to show that  $\text{InF}(\mathcal{E}|\rho)$  does not contain any term linear in  $\beta$ .

The equality  $\text{tr}_B \text{ad}_{\Omega_B} = 0$  from Eq. (14) allows us to freely apply a pure-bath time evolution before  $\text{tr}_B$ :

$$\text{tr}_B \text{Ad} e^{i\alpha\Omega_B} = \text{tr}_B e^{i\alpha \text{ad}_{\Omega_B}} = \text{tr}_B e^0 = \text{tr}_B, \quad (\text{A1})$$

where  $\text{Ad } X(\cdot) \equiv X(\cdot)X^\dagger$ . Therefore

$$\begin{aligned} \mathcal{E}(\rho_S) &= \text{tr}_B \text{Ad} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} \rho(0) \\ &= \text{tr}_B \text{Ad} e^{i\alpha\Omega_B} \text{Ad} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} \rho(0) \quad (\text{A2}) \\ &= \text{tr}_B \text{Ad} e^{i\alpha\Omega_B} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} \rho(0). \end{aligned}$$

We can expand the product of the above exponential functions as a Taylor series of  $\beta$ :

$$\begin{aligned} e^{i\alpha\Omega_B} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} &= 1 + \beta e^{i\alpha\Omega_B} \frac{d}{d\beta} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} \Big|_{\beta=0} \\ &\quad + \frac{\beta^2}{2} e^{i\alpha\Omega_B} \frac{d^2}{d\beta^2} e^{-i(\alpha\Omega_B + \beta\Omega_{SB})} \Big|_{\beta=0} + \mathcal{O}(\beta^3) \\ &= 1 - i\beta K_1 - \frac{\beta^2}{2}(K_1^2 + iK_2) + \mathcal{O}(\beta^3), \quad (\text{A3}) \end{aligned}$$

where we have defined two Hermitian operators,

$$K_1 = \sum_{n=0}^{\infty} \frac{(i\alpha)^n}{(n+1)!} \text{ad}_{\Omega_B}^n(\Omega_{SB}), \quad (\text{A4})$$

$$K_2 = i \sum_{n=1}^{\infty} \frac{(i\alpha)^n}{(n+2)!} \sum_{k=1}^n \text{ad}_{\Omega_B}^{n-k} \text{ad}_{\Omega_{SB}} \text{ad}_{\Omega_B}^k(\Omega_{SB}) \quad (\text{A5})$$

According to the exponential function derivative formula:

$$\frac{d}{dt} e^{\Omega(t)} = e^{\Omega(t)} \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)!} \text{ad}_{\Omega(t)}^n \left( \frac{d}{dt} \Omega(t) \right). \quad (\text{A6})$$

Apply the Taylor series to Eq. (A2) and expand the conjugation, we have

$$\begin{aligned} \mathcal{E}(\rho_S) &= \rho_S - i\beta \text{tr}_B[K_1, \rho(0)] - i\frac{\beta^2}{2} \text{tr}_B[K_2, \rho(0)] \\ &\quad - \frac{\beta^2}{2} \text{tr}_B[K_1, [K_1, \rho(0)]] + \mathcal{O}(\beta^3). \quad (\text{A7}) \end{aligned}$$

Now that the second and third term have the single commutator form, which vanishes after taking trace product with  $\rho_S$  (see Eq. (13)). The smallest term left in the infidelity expression is therefore the  $\beta^2$  term with conjugation by  $K_1(\alpha)$ . Thus proves our proposition.

### Appendix B: Norm inequalities

### Appendix C: Leading orders of polynomials

To make the idea of “leading power” precise, we take the following conventions. For a polynomial  $\sum_k a_k x^k$  with real coefficients, we define its leading power as the smallest number  $k$  such that  $a_k \neq 0$ . For a polynomial with multiple parameters, the leading power is defined as the smallest combined power with nonzero coefficient. We say  $p \sim q$  if  $p$  and  $q$  have the same leading power,  $p \ll q$  if the leading power of  $p$  is greater than  $q$ ,  $p \lesssim q$  if the leading power of  $p$  is greater or equal to  $q$ . This establishes a transitive ordering for the polynomials. This notation is consistent with the usual inequalities for non-negative polynomials (i.e., polynomials with nonnegative output for nonnegative input parameters). If two non-negative polynomial satisfies  $p \leq q$  for all nonnegative parameters, then  $p \lesssim q$ ; but the converse cannot be said. If  $p \lesssim q$ , then  $p + q \sim q$ . In general  $\sum_i p_i \sim \max_i\{p_i\}$ , where the maximization is defined by  $p_i \lesssim \max_i\{p_i\}, \forall i$  and represents the leading order term of the series  $\sum_i p_i$ .

### Appendix D: Implementation for channel infidelity and channel trace distance

For a single-qubit system coupled with a bath, we denote the free evolution Hamiltonian for time  $\tau_0$  as:

$$\Omega = \tau_0 H = \sigma_0 \otimes B_0 + \sum_{i=1}^3 \sigma_i \otimes B_i, \quad (\text{D1})$$

where  $\{\sigma_i\}$  are the system space Pauli operators and  $\{B_i\}$  are the Hermitian bath operators with  $\text{tr}(B_0) = 0$ . The bath operators  $B_i$  can be arbitrary dimensional, but are normalized such that

$$\|\Omega\| = \left( \left\| \sum_{i=1}^3 \sigma_i \otimes B_i \right\| \right) \equiv \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (\text{D2})$$

To reduce noise, we perform dynamical decoupling by periodically applying the pulse cycle “ $-X - Z - X - Z$ ”. This is also known as PDD scheme. This cycle is sometimes referred to the universal decoupling sequence and it can suppress all types of noise on the qubit. Under the ideal-pulse assumption, the time evolution operator for the PDD cycle is:

$$Z e^{-i\Omega} X e^{-i\Omega} Z e^{-i\Omega} X e^{-i\Omega} \equiv e^{-i\Omega_p}, \quad (\text{D3})$$

where an effective Hamiltonian  $\Omega_p$  is defined for the time evolution. When the pulses are imperfect, we model the noisy in the pulses with a noise operator  $V(\eta)$  such that the pulses are given by  $\tilde{P}_i = P_i V(\eta)$  where  $\eta$  is a generic parameter to characterize the noise strength. It follows the time evolution operator for the noisy PDD cycle is given by

$$\tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \tilde{Z} e^{-i\Omega} \tilde{X} e^{-i\Omega} \equiv e^{-i\tilde{\Omega}_p}, \quad (\text{D4})$$

In the following, we consider the unitary noise, often due to over rotation, e.g.,  $\tilde{P}_i = P_i e^{-i\vec{\sigma} \cdot \vec{\theta}}$ , where  $\vec{\theta} = (\theta_1, \theta_2, \theta_3)^\top$ . In this case, the noise strength parameter  $\eta$  is characterized by  $\eta \equiv \|\vec{\theta}\| \equiv \theta$ .

The output system state under the free evolution Hamiltonian  $\Omega$  and the PDD effective Hamiltonian  $\Omega_P$  given an arbitrary initial state  $\rho_S \in \mathcal{H}_S$  and can be written as

$$\mathcal{E}_\Omega(\rho_S) = \text{tr}_B(e^{-i\Omega}(\rho_S \otimes \rho_B)e^{i\Omega}), \quad (\text{D5})$$

$$\mathcal{E}_{\Omega_P}(\rho_S) = \text{tr}_B(e^{-i\Omega_P}(\rho_S \otimes \rho_B)e^{i\Omega_P}), \quad (\text{D6})$$

where  $\rho_B \in \mathcal{H}_B$  is the initial bath state;  $\text{tr}_B(\cdot)$  is the partial trace over  $\mathcal{H}_B$ . Since the noise map  $\mathcal{E}$  is determined by  $\Omega$  and  $\rho_B$ , to evaluate the channel infidelity, we consider the infinite temperature state (maximally mixed state) to be  $\rho_B$  and random bath operators  $\{B_i\}$  for both  $\Omega$  and  $\Omega_P$ .

Given a channel  $\mathcal{E}$ , the trace distance and infidelity are given by

$$D(\mathcal{E}|\rho) \equiv D(\rho, \mathcal{E}(\rho)) = \frac{1}{2}\|(\mathcal{I} - \mathcal{E})(\rho)\|_1, \quad (\text{D7})$$

$$\text{Inf}(\mathcal{E}|\rho) \equiv 1 - F(\rho, \mathcal{E}(\rho)) = \text{tr}(\rho(\mathcal{I} - \mathcal{E})(\rho)). \quad (\text{D8})$$

respectively. To get rid of the state dependence and Hamiltonian dependence (due to random bath operators), we consider a maximization of the trace distance and infidelity over all the pure input state and random bath operators,

$$D_{\max} = \max_{\mathcal{E}, \rho \in \mathcal{D}(\mathcal{H})} \frac{1}{2}\|(\mathcal{I} - \mathcal{E})(\rho)\|_1 \equiv \frac{1}{2}\|\mathcal{I} - \mathcal{E}\|_1, \quad (\text{D9})$$

$$\text{Inf}_{\max} \equiv \max_{\mathcal{E}, \rho \in \mathcal{D}(\mathcal{H})} \text{tr}(\rho(\mathcal{I} - \mathcal{E})(\rho)). \quad (\text{D10})$$

We restrict ourselves to the infidelity of PDD channel and free evolution channel, namely  $\text{Inf}_{\max}^{\text{DD}}(\beta, \alpha, \eta)$ ,  $\text{Inf}_{\max}^{\text{free}}(\beta, \alpha)$ .

In the figures, we thus consider the ratio of the channel infidelity and channel trace distance for DD sequence and that for free evolution, i.e.,  $\text{Inf}_{\max}^{\text{DD}} / \text{Inf}_{\max}^{\text{free}}$  and  $D_{\max}^{\text{DD}} / D_{\max}^{\text{free}}$ , respectively. The maximization is done over 10000 samples for different initials states and random bath operators with a fixed bath density matrix given by the infinite temperature state (maximally mixed state).