Combining dynamical decoupling with fault-tolerant quantum computation

Hui Khoon Ng,[1,*] Daniel A. Lidar,[2] and John Preskill[1]

[1]Institute for Quantum Information, California Institute of Technology, Pasadena, CA 91125, USA
[2]Departments of Chemistry, Electrical Engineering, and Physics,
and Center for Quantum Information Science & Technology,
University of Southern California, Los Angeles, California 90089, USA

We study how dynamical decoupling (DD) pulse sequences can improve the reliability of quantum computers. We prove upper bounds on the accuracy of DD-protected quantum gates and derive sufficient conditions for DD-protected gates to outperform unprotected gates. Under suitable conditions, fault-tolerant quantum circuits constructed from DD-protected gates can tolerate stronger noise, and have a lower overhead cost, than fault-tolerant circuits constructed from unprotected gates. Our accuracy estimates depend on the dynamics of the bath that couples to the quantum computer, and can be expressed either in terms of the operator norm of the bath's Hamiltonian or in terms of the power spectrum of bath correlations; we explain in particular how the performance of recursively generated concatenated pulse sequences can be analyzed from either viewpoint. Our results apply to Hamiltonian noise models with limited spatial correlations.

PACS numbers: 03.67.Pp, 03.67.Lx

I. INTRODUCTION

Two well-known methods for protecting quantum systems from noise are dynamical decoupling (DD) and quantum error correction (QEC). In DD, pulses are applied to the protected system, chosen so that the damaging effects of the noise nearly average away. In QEC, protected logical qubits are encoded as collective states of many physical qubits, chosen so that damage due to noise can be detected and reversed.

Each method has advantages and disadvantages. On the plus side, resource requirements for DD are relatively modest [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Only unitary control operations need be applied to the system; there is no need to perform measurements or to replace used ancillary qubits with fresh qubits. Furthermore, a single physical qubit suffices for each protected logical qubit, and protected quantum gates can be implemented using relatively short sequences of pulses. DD pulse sequences are simple enough that experiments on a wide variety of quantum systems have convincingly demonstrated the effectiveness of DD [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. On the minus side, DD is effective only against low-frequency noise, slowly varying on the time scale set by the interval between pulses, and its effectiveness is intrinsically limited by imperfections in the timing and shape of the pulses. Furthermore, DD is not an efficient scheme for flushing entropy from the system, if no qubits are replaced or refreshed; thus it seems that DD does not by itself provide a feasible route to scalable quantum computing.

On the plus side, the quantum accuracy threshold theorem establishes that QEC, through judicious design of fault-tolerant gadgets acting on code blocks, suffices for accurate simulation of arbitrarily long quantum computations, if the noise is sufficiently weak and reasonably local [26, 27, 28, 29, 30, 31, 32, 33]. QEC can succeed against high-frequency noise, where DD methods fail. On the minus side, though, the resource requirements for QEC are quite daunting. A ready supply of fresh qubits is necessary; furthermore, the number of physical system qubits needed to encode one logical qubit, and the number of physical gates needed to execute one logical gate, can be substantial. Because of the complexity of fault-tolerant quantum computing protocols, and because these protocols work only when the noise is already quite weak, experiments showing that QEC can suppress naturally occurring noise have not yet been performed.

Because of their complementary strengths, DD and QEC used together should be more effective at protecting quantum computers from noise than either used by itself. Combining these two methods of error control is the topic of this paper. Hybrid schemes combining DD with QEC have been proposed previously [34, 35, 36], and even studied experimentally [37]. Our new contribution is a systematic investigation of the advantages of hybrid schemes for fault-tolerant quantum computing, including rigorous bounds on performance.

Our main technical results are analytic expressions for the "effective noise strength" of quantum gates implemented using DD pulse sequences. The effective noise strength is (an upper bound on) the deviation in the operator norm of the noisy protected gate from an ideal gate. In the Hamiltonian noise models that we consider, the logarithm of the operator realized by a DD-protected gate can be expanded as a power series (the Magnus expansion) in the noise Hamiltonian; we derive upper bounds on the sum of this series, obtaining formulas for the effective noise strength in terms of parameters in the noise Hamiltonian. We find such bounds both for general DD pulse sequences, and also for pulse sequences that

*Current address: DSO National Laboratories, Applied Physics Lab, 20 Science Park Drive, Singapore 118230, Singapore

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **NOV 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE **Combining dynamical decoupling with fault-tolerant quantum computation**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **California Institute of Technology,Institute for Quantum Information,Pasadena,CA,91125**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Report number: CALT-68-2753

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**39**	

have an approximate time-reversal symmetry; in the latter case the terms of even order in the Magnus expansion are heavily suppressed.

Armed with our formulas for the effective noise strength, we can derive a "noise-suppression threshold condition" on the noise parameters. When this condition is satisfied, DD-protected gates are more accurate than unprotected gates. We can also compare fault-tolerant quantum circuits composed from DD-protected gates with circuits composed from unprotected gates. In either case, we can express the "accuracy threshold condition" on the noise parameters. When this condition is satisfied, quantum computation is scalable — accurate computations of arbitrary size can be performed with a reasonable overhead cost. Typically, improvements in gate accuracy achieved by DD mean that more noise can be tolerated by QEC combined with DD than by QEC alone, and that invoking DD can reduce the overhead cost of QEC.

Our expressions based on the Magnus expansion for the effective noise strength depend on the operator norm of the Hamiltonian that governs the internal dynamics of the quantum computer's environment (the "bath"), and the results are not useful if this norm is large. But we also describe an alternative method of analysis yielding expressions for the effective noise strength in terms of the frequency spectrum of the bath correlations. Results derived by this method can be applicable even if the bath Hamiltonian has unbounded norm, as long as the typical bath frequencies are sufficiently small.

The performance of DD can sometimes be enhanced by using recursively generated "concatenated" pulse sequences [10]. Adding an extra "level" to the recursive hierarchy further suppresses the effective noise Hamiltonian, but at the cost of lengthening the pulse sequence, and the minimal effective noise strength is achieved by choosing the level that optimizes this tradeoff. We analyze concatenated DD sequences and estimate the optimal effective noise strength, using both our bounds on the Magnus expansion and the correlation function viewpoint.

Our analysis of the improvement in gate accuracy that can be achieved by combining DD and QEC applies only to a special class of Hamiltonian noise models. These models satisfy what we call the "local-bath assumption" which limits the spatial correlations in the noise. Whether our results can be extended to more general noise models that violate this assumption is an intriguing open question.

We formulate our noise model in Sec. II. In Sec. III and Sec. IV we review and develop some of the tools we need to analyze the performance of DD pulse sequences and fault-tolerant quantum circuits using the Magnus expansion. We state our central results relating the effective noise strength of DD-protected gates to the properties of the noise Hamiltonian, and their implications concerning the noise-suppression threshold and accuracy threshold, in Sec. V; then we apply these results to some specific pulse sequences in Sec. VI. Derivations of these results are contained in Sec. VII and the Appendices. We analyze concatenated DD in Sec. VIII. In Sec. IX, we emphasize that the effective noise can be related to intensive quantities that are independent of the spatial volume of the bath, and in Sec. X we express the noise strength in terms of properties of bath correlations. Sec. XI contains our conclusions.

II. NOISE MODEL

A. Noise Hamiltonian

We denote by S the system consisting of all of the qubits in our quantum computer, and we describe the noise acting on S using a "noise Hamiltonian" H, which governs the joint evolution of the system and its environment, the bath B. During a computation, the Hamiltonian also contains time-dependent terms that realize quantum gates acting on the qubits, but for now consider the case where there are no gates; then H may be expressed as

$$H \equiv H_B + H_{\rm err}. \tag{1}$$

Here

$$H_B \equiv \mathbb{I}_S \otimes B_0 \tag{2}$$

describes the "free" evolution of the bath (how it would evolve if it were not coupled to the system) while

$$H_{\rm err} \equiv H_S^0 + H_{SB} \tag{3}$$

includes all the terms in H that act non-trivially on the system. The term

$$H_S^0 \equiv S_0 \otimes \mathbb{I}_B \tag{4}$$

describes the unperturbed free evolution of the system; H_{SB} contains terms coupling the system to the bath, and also perhaps other noise terms that act nontrivially only on the system.

Though for some purposes it may seem natural to transform away H_S^0 by working in the interaction picture (that is, by considering the motion of the system relative to the rotating frame determined by H_S^0), we have included H_S^0 in the term $H_{\rm err}$ that represents the noise acting on the system. Our reason is that the DD sequences we study are designed to remove the effects of all "always-on" terms in the Hamiltonian that act on the system, *i.e.*, not just H_{SB} but also the free evolution term H_S^0. We may exclude from the noise Hamiltonian H a term proportional to $\mathbb{I}_S \otimes \mathbb{I}_B$, which just shifts the zero point of the energy and has no dynamical effect. Alternatively, we can choose the coefficient of this term to optimize our estimate of the accuracy threshold; for example, we may by convention choose H to be traceless.

Now consider modeling the noise during a nontrivial quantum computation. A computation is a circuit containing three types of operations: qubit state preparations, unitary quantum gates, and qubit measurements. We model a noisy preparation as an ideal preparation followed by evolution according to H for a specified time interval, and we model noisy measurements as ideal measurements preceded by evolution according to H. We assume that quantum gates are executed using short, hard pulses, where, as in some experiments, the time interval between consecutive pulses is much longer than the pulse width. Each pulse has its support in a narrow interval of width δ, and we denote by τ_0 the sum of the pulse width and the pulse interval (see Fig. 1), where $\delta \ll \tau_0$. To be concrete, we will sometimes assume that the pulses are perfectly "rectangular" — *i.e.*, have vanishing rise-time and fall-time. However, the details of the pulse shape are not really used in our analysis; rather, all that matters is that the pulse is confined to a narrow interval (and even this assumption will be relaxed in Sec. VI C). We use the same noise Hamiltonian H to describe the noise both during a pulse and during the interval between pulses. We neglect errors in the timing and strength of the pulses; these are typically small in practice because the pulses are controlled by accurate classical circuitry.

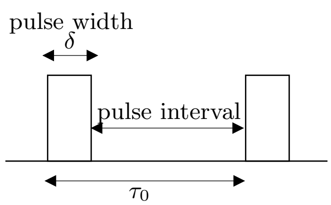

FIG. 1: Parameters characterizing a sequence of uniformly spaced rectangular pulses: δ is the pulse width, and $\tau_0 - \delta$ is the interval between the end of one pulse and the beginning of the following pulse.

B. Local-bath assumption

To further simplify our analysis, we make an additional assumption about the noise, which we call the *local-bath assumption* [30], illustrated in Fig. 2. Let us use the term "location" to speak of an operation in a quantum circuit that is performed in a single time step — a location may be a single-qubit or multi-qubit gate, a qubit preparation step, a qubit measurement, or the identity operation in the case of a qubit that is idle during the time step. Each time step has duration t_0; thus $t_0 = N\tau_0$ if N equally spaced pulses are applied at a particular location. For a specified location labeled by a, let $\mathcal{Q}_a$ denote the set of qubits that participate in the operation applied at that location (for example, a pair of qubits if the operation is a two-qubit gate). Under the local-bath assumption, the noise Hamiltonian can be expressed as

$$H = \sum_a H_a, \tag{5}$$

where the sum is over all locations occurring at a particular time step, and where for any two distinct locations a and b in that time step, H_a and H_b act not only on disjoint sets of system qubits but also on distinct baths. That is, we may write

$$H_a = H_{B,a} + H_{\mathrm{err},a}, \tag{6}$$

with

$$\begin{aligned} H_{B,a} &= \mathbb{I}_{S,a} \otimes B_{0,a}, \\ H_{\mathrm{err},a} &= \sum_\alpha S_{\alpha,a} \otimes B_{\alpha,a}, \end{aligned} \tag{7}$$

where the operators $S_{\alpha,a}$ act on $\mathcal{Q}_a$, and where, for $a \neq b$, the bath operators $B_{0,a}$ and $B_{\alpha,a}$ associated with location a commute with the bath operators $B_{0,b}$ and $B_{\alpha,b}$ associated with location b. Each H_a is assumed to be time-independent during the duration of location a (this assumption is helpful because DD pulse sequences are typically designed to cope with a time-independent noise Hamiltonian), but Hamiltonians at different locations need not be the same.

The local-bath assumption allows us express the time evolution operator for a single time step as a product of unitary operators, each associated with one particular location, and to analyze the effectiveness of the DD pulse sequence for each location separately. Though our local-bath model may be a reasonable approximation to the noise in actual systems, interactions between qubits (and their associated baths) at different locations are surely present at some level, and in Sec. IX we will comment further on how our analysis is affected when the local-bath assumption is relaxed.

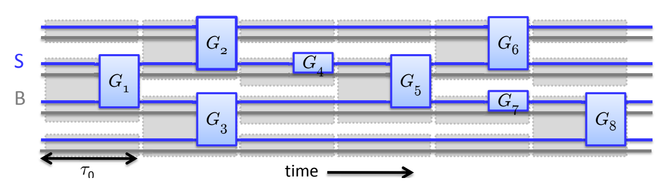

FIG. 2: Illustration of the local-bath assumption. Blue lines are system qubits and grey lines are bath subsystems. Each blue rectangle is a quantum gate, and its associated light grey rectangle represents the joint evolution of system qubits and bath subsystems that are coupled by the noise Hamiltonian. The light grey rectangles do not overlap, indicating that gates executed in parallel act not only on distinct system qubits but also on distinct bath subsystems.

C. Noise parameters

To characterize the noise strength, it is useful to introduce the parameters β, J, ϵ defined by:

$$\beta \equiv \max_a \|H_{B,a}\|, \tag{8}$$

$$J \equiv \max_a \|H_{\mathrm{err},a}\|, \tag{9}$$

$$\epsilon \equiv \beta + J \geq \max_a \|H_a\|. \tag{10}$$

The norm here is the sup operator norm

$$\|A\| \equiv \sup_{\{|v\rangle\}} \frac{\|A|v\rangle\|}{\||v\rangle\|}, \tag{11}$$

where the vector norm is the Euclidean norm $\||v\rangle\| \equiv \sqrt{\langle v|v\rangle}$. Actually, our results concerning the effectiveness of DD pulse sequences apply for any norm that is unitarily-invariant (and therefore also submultiplicative [38]), but the operator norm will be used to relate these results to the accuracy threshold for fault-tolerant quantum computing [30, 31]. We are typically interested in the case where the noise is weak, in the sense that the dimensionless parameter $\epsilon\tau_0$ is small compared to one (and hence also $J\tau_0 \ll 1$ and $\beta\tau_0 \ll 1$). We will derive bounds on the performance of DD-protected quantum gates expressed in terms of these small quantities, and also in terms of the dimensionless pulse width $\delta/\tau_0 \ll 1$.

For our analysis of fault-tolerant circuits, we will find it convenient to assume that measurements and preparations are at least as fast as pulses, *i.e.*, can be executed in time at most δ. But in Sec. V D we will discuss how to interpret our results if measurements or preparations take much longer than pulses.

III. TOOLS

Let us next introduce some tools for analyzing the noise suppression arising from DD techniques. We focus here on the foundations of our analysis based on the Magnus expansion; further background, needed for our analysis based on bath correlation functions, will be discussed in Sec. X. We also provide here a brief discussion of fault tolerance, including the notion of the effective noise strength at a circuit location, a central quantity in our analysis.

A. Toggling frame

For now, disregard that we want to do computation, and focus instead on quantum storage — the original context for DD methods. In the absence of any external control, the system and bath evolve under the time-independent noise Hamiltonian H. A DD pulse sequence is realized via a time-dependent control Hamiltonian $H_c(t)$ acting only on the system so that the system and bath evolve according to $H + H_c(t)$. (In our noise model, we assume that the same noise Hamiltonian H applies during a pulse as between pulses, while recognizing that this assumption is really an idealization.) The DD sequence can be described using either $H_c(t)$ itself or using the time evolution operator $U_c(t) \equiv U_c(t,0)$ generated by $H_c(t)$.

For understanding the effects of the control Hamiltonian, it is convenient to use the interaction picture defined by $H_c(t)$, also known as the *toggling frame* [1]. The toggling-frame density operator $\tilde{\rho}_{SB}(t)$ is related to the Schrödinger-picture density operator $\rho_{SB}(t)$ by

$$\begin{aligned}\rho_{SB}(t) &= U(t,0)\rho_{SB}(0)U^\dagger(t,0)\\ &\equiv U_c(t)\tilde{\rho}_{SB}(t)U_c^\dagger(t),\end{aligned} \tag{12}$$

where $U(t,0)$ is the evolution operator generated by the full Hamiltonian $H+H_c(t)$. Therefore the toggling-frame state evolves according to

$$\tilde{\rho}_{SB}(t) = \tilde{U}(t,0)\tilde{\rho}_{SB}(0)\tilde{U}^\dagger(t,0), \tag{13}$$

where the toggling-frame time evolution operator

$$\tilde{U}(t,0) \equiv U_c^\dagger(t)U(t,0) \tag{14}$$

is generated by the *toggling-frame Hamiltonian*

$$\tilde{H}(t) \equiv U_c^\dagger(t)HU_c(t). \tag{15}$$

Since $U_c(t)$ acts nontrivially only on the system, $\tilde{H}(t)$ can be written as

$$\tilde{H}(t) = H_B + \tilde{H}_{\mathrm{err}}(t), \tag{16}$$

where $\tilde{H}_{\mathrm{err}}(t) \equiv U_c^\dagger(t)H_{\mathrm{err}}U_c(t)$ is the toggling-frame version of H_{err}. Because the operator norm is unitarily-invariant, we have $\|\tilde{H}(t)\| = \|H\| \leq \epsilon$ and $\|\tilde{H}_{\mathrm{err}}(t)\| = \|H_{\mathrm{err}}\| \leq J$.

We consider *cyclic DD*, where $U_c(t)$ returns to the identity (up to a possible irrelevant overall phase) at the end of a cycle taking time t_{DD}:

$$U_c(t_{\mathrm{DD}}) = U_c(0) = \mathbb{I}. \tag{17}$$

Therefore, at the end of the cycle, the toggling-frame and Schrödinger-picture states coincide.

B. Finite-width pulses

In DD, the system is controlled using a sequence of pulses, where the control Hamiltonian $H_c(t)$ vanishes in between the pulses. The control unitary resulting from a sequence of R pulses can be expressed as

$$U_c = \mathbb{I}P_R\mathbb{I}P_{R-1}\mathbb{I}\ldots P_2\mathbb{I}P_1\mathbb{I}. \tag{18}$$

where P_k is the unitary achieved by the kth pulse. We have inserted the identity $\mathbb{I}$ between successive pulses to indicate the time intervals during which $H_c(t) = 0$. For some pulse sequences, including the ones described in Sec. VI, all pulse intervals have the same duration, but for most of our analysis (excluding some of the discussion of pulse-width effects in Sec. VII A 1) we need not assume that the pulses are uniformly spaced. (It is known that the effectiveness of DD can sometimes be improved by varying the spacing between pulses [13, 21, 39, 40, 41, 42].)

If the pulses are rectangular with width δ, then we may write

$$P_k \equiv \exp(-i\delta H_{P_k}), \tag{19}$$

where H_{P_k} is the time-independent control Hamiltonian that is turned on during the kth pulse. If the kth pulse begins at time s_k, then the control unitary during the pulse ($t \in [s_k, s_k + \delta)$) is

$$\begin{aligned} U_c(t) &= \exp\left(-i\Delta_k H_{P_k}\right) U_c(s_k) \\ &= \exp\left(-i\Delta_k H_{P_k}\right) P_{k-1} \ldots P_2 P_1, \end{aligned} \tag{20}$$

where $\Delta_k = t - s_k$. The toggling-frame Hamiltonian $\tilde{H}(t)$ can be written as

$$\begin{aligned} &\tilde{H}(t) = U_c^\dagger(t) H U_c(t) \qquad (21) \\ &= \begin{cases} \tilde{H}^{(k-1)} & \text{for } t \in [s_{k-1} + \delta, s_k), \\ e^{i\Delta_k \tilde{H}_{P_k}^{(k-1)}} \tilde{H}^{(k-1)} e^{-i\Delta_k \tilde{H}_{P_k}^{(k-1)}} & \text{for } t \in [s_k, s_k + \delta), \end{cases} \end{aligned}$$

where

$$\tilde{H}^{(k-1)} = P_1^\dagger P_2^\dagger \ldots P_{k-1}^\dagger H P_{k-1} \ldots P_2 P_1. \tag{22}$$

In the case of cyclic DD, after the last pulse of a complete cycle we have $U_c = \mathbb{I}$ and $\tilde{H} = H$.

C. Magnus expansion

For a given $\tilde{H}(t)$, the toggling-frame time evolution operator $\tilde{U}(t_{\mathrm{DD}}, 0)$ can be computed using a *Magnus expansion* [43]. For a unitary time evolution operator $U_M(t,0)$ satisfying the Schrödinger equation

$$i\frac{\partial}{\partial t} U_M(t,0) = H_M(t) U_M(t,0), \quad U_M(0,0) = \mathbb{I}, \tag{23}$$

determined by Hamiltonian $H_M(t)$, the Magnus expansion at time T is an operator series

$$\Omega(T) \equiv \sum_{n=1}^{\infty} \Omega_n(T) \tag{24}$$

such that

$$U_M(T,0) = \exp\left[\Omega(T)\right], \tag{25}$$

and Ω_n is nth order in the Hamiltonian $H_M(t)$. Thus, for the fixed time T, time evolution generated by the time-dependent Hamiltonian $H_M(t)$ is equivalent to time evolution generated by the time-independent effective Hamiltonian $H_{\mathrm{eff}} \equiv \frac{i}{T}\Omega(T)$.

The leading terms in the Magnus expansion are (see for example, [44])

$$\Omega_1(T) = -i \int_0^T ds\ H_M(s), \tag{26}$$

$$\Omega_2(T) = -\frac{1}{2} \int_0^T ds_1 \int_0^{s_1} ds_2 \left[H_M(s_1), H_M(s_2)\right], \tag{27}$$

$$\begin{aligned} \Omega_3(T) = \frac{i}{6} \int_0^T ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 & \\ \big([H_M(s_1), [H_M(s_2), H_M(s_3)]] & \\ + [H_M(s_3), [H_M(s_2), H_M(s_1)]] \big). & \end{aligned} \tag{28}$$

Higher-order terms can be computed using a recursive formula; see Sec. VII and Appendix A. In general, $\Omega_n(T)$ is the time integral of a sum of $(n-1)$-nested commutators, each with n factors of $H_M(t)$. The Magnus expansion is thus an infinite series in $H_M T$; a sufficient condition for convergence is [45]

$$\int_0^T dt\ \|H_M(t)\| < \pi. \tag{29}$$

For cyclic DD, we can use the Magnus expansion to compute the toggling-frame time evolution operator $\tilde{U}(t_{\mathrm{DD}}, 0)$ for one complete cycle, where $H_M(t)$ is the toggling-frame Hamiltonian $\tilde{H}(t) = U_c^\dagger(t) H U_c(t)$ and $U_c(t_{\mathrm{DD}}) = \mathbb{I}$. In first order we obtain

$$\begin{aligned} \Omega_1(t_{\mathrm{DD}}) &= -i \int_0^{t_{\mathrm{DD}}} dt\ \tilde{H}(t) \\ &= -i \int_0^{t_{\mathrm{DD}}} dt\ U_c^\dagger(t) H U_c(t). \end{aligned} \tag{30}$$

For group-based DD schemes, like the examples we will discuss in Sec. VI, the integral Eq. (30) averages H over a finite group $\mathcal{G}$ if the pulses are ideal, projecting H into the commutant of $\mathcal{G}$ [3]. If $\mathcal{G}$ acts irreducibly on the system Hilbert space, the commutant contains only the identity operator acting on the system, and therefore $\Omega_1(t_{\mathrm{DD}})$ acts nontrivially only on the bath. In that case we say that $\Omega_1(t_{\mathrm{DD}})$ is a "pure bath" term.

We say that a DD pulse sequence achieves first-order decoupling if the first-order term in the Magnus expansion for $U_c(t_{\mathrm{DD}})$ acts trivially on the system. More generally, the sequence achieves mth-order decoupling if $\Omega_n(t_{\mathrm{DD}})$ is a pure bath term for each $n \leq m$. In our analysis we will at first consider pulse sequences that achieve first-order decoupling for ideal zero-width pulses (later we will discuss the corrections to first-order decoupling that arise when the pulses have nonzero width, and we will also describe "Eulerian" pulse sequences that achieve first-order decoupling even when pulse widths are nonzero). In particular, these pulse sequences have the property

$$\int_0^{t_{\rm DD}} dt\ \tilde{H}_{\rm err,0}(t) = 0, \tag{31}$$

where the subscript "0" on $\tilde{H}_{\rm err,0}(t)$ indicates that the toggling-frame Hamiltonian $\tilde{H}_{\rm err}(t)$ is considered in the limit $\delta \to 0$, while holding τ_0 fixed. (Eq. (31) implies that there is no term in $H_{\rm err,0}$ that acts nontrivially on the system and commutes with $H_c(t)$ for all $t \in [0, t_{\rm DD}]$; such terms would not be removed by the DD sequence described by $H_c(t)$.) For pulse sequences satisfying Eq. (31) it follows from Eq. (16) that the first-order term in the Magnus expansion is

$$\begin{aligned}\Omega_1(t_{\rm DD}) &= -i\int_0^{t_{\rm DD}} dt\ \tilde{H}_0(t)\\ &= -iH_B t_{\rm DD} - i\int_0^{t_{\rm DD}} dt\ \tilde{H}_{\rm err,0}(t)\\ &= -iH_B t_{\rm DD},\end{aligned} \tag{32}$$

a pure bath term, when $\delta = 0$. For pulses with nonzero width δ, first-order decoupling is not exact, but the deviation of $\Omega_1(t_{\rm DD})$ from a pure bath term is $O(\delta/\tau_0)$ and thus small when the pulses are sufficiently narrow. For suitably designed pulse sequences the deviation can be improved to a higher power of δ/τ_0 [7, 46].

A pulse sequence that achieves first-order decoupling will also achieve second-order decoupling if $\tilde{H}$ is time-symmetric: $\tilde{H}(t_{\rm DD} - t) = \tilde{H}(t)$ for $t \in [0, t_{\rm DD}]$. This condition is satisfied provided

$$U_c(t_{\rm DD} - t) = V_t U_c(t), \tag{33}$$

where V_t is unitary and commutes with H (for example, if $V_t = e^{i\phi_t}\mathbb{I}$ is a phase). In fact, when $\tilde{H}$ is time-symmetric, not just the second-order term, but all even terms in the Magnus expansion vanish, as we show in Appendix B.

D. Quantum accuracy threshold theorem

The quantum accuracy threshold theorem establishes that a noisy quantum computer can operate reliably if the noise is sufficiently weak. Under the local-bath assumption formulated in Sec. II, the operation applied at location a in the noisy circuit is a unitary transformation $\overline{G}_a$ acting on the system and bath, which can be expressed as

$$\overline{G}_a = \mathcal{G}_a + \mathcal{B}_a. \tag{34}$$

Here $\mathcal{G}_a$ is the "good part" of the operation; it can be expressed as $G_a \otimes B_a$, where G_a is the ideal operation that would be applied to the system in the absence of noise, and B_a is a unitary transformation acting on the bath. The operator $\mathcal{B}_a$ is the "bad part," the deviation of $\overline{G}_a$ from the ideal operation, which acts jointly on system and bath. (Recall that we model a noisy qubit preparation as an ideal preparation followed by a noisy unitary transformation, and a noisy qubit measurement as a noisy unitary transformation followed by an ideal measurement; for preparation or measurement locations, $\overline{G}_a$ denotes the noisy transformation that follows or precedes the ideal preparation or measurement.) In this noise model, we may characterize the noise strength by

$$\bar{\eta} \equiv \max_a \|\mathcal{B}_a\|, \tag{35}$$

the maximum value of the operator norm of the bad part, where the maximum is with respect to all locations in the noisy circuit. The threshold theorem asserts that an ideal circuit of arbitrary size can be simulated accurately if $\bar{\eta}$ is less than a critical value η_0, the accuracy threshold. The threshold theorem proved in [31] actually applies to a broader class of noise models that do not necessarily satisfy the local-bath assumption, but this class includes the noise model of Sec. II as a special case. The analysis in [47] established a lower bound on the accuracy threshold, $\eta_0 \gtrsim 10^{-4}$.

In this paper we will relate the noise strength $\bar{\eta}$ defined by Eq. (34) and Eq. (35) to the parameters that characterize the noise model defined in Sec. II. We denote by $\eta_{\rm DD}$ the value of $\bar{\eta}$ that can be achieved using dynamical decoupling, and we denote by η the value of $\bar{\eta}$ achieved without using dynamical decoupling. If $\eta_{\rm DD} < \eta$, then we say that the noise is below the *noise suppression threshold*, meaning that dynamical decoupling reduces the effective noise strength. If $\bar{\eta} < \eta_0$, then we say the noise is below the accuracy threshold, meaning that scalable quantum computing is possible.

In Sec. V we express $\eta_{\rm DD}$ in terms of the parameters J and ϵ defined in Eq. (8)-(10). In Sec. X we express $\eta_{\rm DD}$ in terms of properties of bath correlation functions, using a different formula than Eq. (35) which is explained in Appendix J.

IV. DD-PROTECTED GATES

A. Including the gate pulse

So far we have described how to reduce the noise in a quantum memory using cyclic DD. Now we want to estimate the effective noise strength achieved by DD for operations other than the identity, so we must explain how DD is used to suppress the noise in these nontrivial operations, We will describe nontrivial quantum gates, postponing discussion of preparations and measurements until later.

We refer to one cycle of the DD pulse sequence for the identity gate as the "memory" sequence. To perform a DD-protected nontrivial gate G_a, we must modify the memory sequence accordingly. In fact our DD pulse sequence for the gate is exactly the same as the memory sequence, except for the very last pulse. If the memory sequence of R pulses ends with a period of trivial evolution, then we *append* a pulse implementing G_a to the end the memory sequence. Thus, if the memory sequence lasts time $t_{\rm DD}$ and the pulse width is δ, then the G_a pulse sequence lasts time $t_0 = t_{\rm DD}+\delta$ and uses $N = R+1$ pulses. If on the other hand the R-pulse memory sequence ends with a nontrivial pulse implementing P, then we combine this pulse and the gate pulse into a single pulse implementing G_aP. Again, we denote the total time for the G_a pulse sequence by t_0, and the total number of pulses by $N(= R)$.

While we assume for simplicity that every pulse has the same width δ, we recognize that in some cases different types of pulses may have different time scales. For example, in recent experiments with quantum dot qubits, X gates are implemented using (fast) exchange couplings and Z gates are implemented using (slow) magnetic field gradients [17]. One may interpret δ as the duration of the longest pulse used, or one could easily refine our analysis by allowing different pulses to have different widths.

In a DD-protected circuit, each G_a gate is replaced by the corresponding DD-protected gate; under the local-bath assumption, the noisy DD-protected gate is a unitary transformation denoted $\overline{G}_a$ acting jointly on the systems and the bath. Though the duration t_0 of a DD-protected gate is longer than the duration τ_0 of an unprotected gate, the DD-protected gate may be more accurate than the unprotected gate, if the noise is weak enough.

At the end of the complete G_a pulse sequence, the unitary operator $U_{c,a}(t_0)$ generated by the control Hamiltonian $H_c(t)$ (which now includes the gate pulse) is

$$U_{c,a}(t_0) = G_a U_c(t_{\rm DD}) = G_a, \tag{36}$$

because the cyclic memory sequence satisfies $U_c(t_{\rm DD}) = \mathbb{I}$ (up to a possible phase). Therefore the noisy DD-protected gate at location a is

$$\overline{G}_a \equiv U_a(t_0, 0) = U_{c,a}(t_0)\tilde{U}_a(t_0, 0) = G_a\tilde{U}_a(t_0, 0), \tag{37}$$

where $\tilde{U}_a(t_0, 0)$ is the toggling-frame time evolution operator. The corresponding toggling-frame Hamiltonian is similar to the toggling-frame Hamiltonian Eq. (21) for the memory sequence, except for the appended gate pulse:

$$\begin{aligned}\tilde{H}_a(t) &= U^\dagger_{c,a}(t) H_a U_{c,a}(t) \qquad\qquad (38)\\ &= \begin{cases} e^{i\Delta_k \tilde{H}^{(k-1)}_{P_k}}\,\tilde{H}^{(k-1)}_a e^{-i\Delta_k \tilde{H}^{(k-1)}_{P_k}} & \text{for } t \in [s_k, s_k+\delta),\\ \tilde{H}^{(k)}_a & \text{for } t \in [s_k+\delta, s_{k+1}),\\ e^{i\Delta_{R+1} \tilde{H}^{(R)}_{G_a}}\,\tilde{H}^{(R)}_a e^{-i\Delta_{R+1} \tilde{H}^{(R)}_{G_a}} & \text{for } t \in [s_{R+1}, s_{R+1}+\delta),\\ G_a^\dagger H_a G_a & \text{for } t = t_0.\end{cases}\end{aligned}$$

Eq. (38) applies to the case where the gate pulse is appended to the end of the memory sequence; the memory sequence contains R equally spaced pulses labeled by $k = 1, 2, \ldots, R$, and the gate pulse begins at time s_{R+1}.

The DD-protected qubit measurement is the memory pulse sequence followed by an ideal measurement. We assume that the measurement takes time δ, the same as the pulse width, so that the duration t_0 of the protected measurement matches the duration of the DD-protected gate. Similarly, the DD-protected qubit preparation is an ideal preparation lasting time δ followed by the memory pulse sequence. See Sec. V D for discussion of how our analysis is modified when preparations and measurements are slow compared to other operations.

B. Effective noise strength

To define the effective noise strength for the DD-protected gate, we divide the noisy gate into a good part and a bad part as in Eq. (34), obtaining

$$\overline{G}_a = \underbrace{G_aU_{B,a}(t_0)}_{\equiv \mathcal{G}_a} + \underbrace{\overline{G}_a - G_aU_{B,a}(t_0)}_{\equiv \mathcal{B}_a} \tag{39}$$

The good part $\mathcal{G}_a$ describes the ideal evolution in the absence of noise ($H_{\rm err} = 0$) — the ideal gate G_a is applied to the system, while the bath evolves according to its unperturbed Hamiltonian $H_{B,a}$. The bad part $\mathcal{B}_a$ describes the effects of noise, as modified by the DD pulse sequence.

Using Eq. (37) and the unitary invariance of the operator norm, we obtain an expression for the noise strength of the DD-protected circuit:

$$\eta_{\rm DD} \equiv \max_a \|\mathcal{B}_a\| = \max_a \left\|\tilde{U}_a(t_0, 0) - U_{B,a}(t_0)\right\|; \tag{40}$$

this is just the norm of the bad part expressed in the toggling frame. In what follows, we will sometimes drop the subscript a when context makes the intended meaning clear.

We can now estimate $\eta_{\rm DD}$ using the Magnus expansion. We write

$$\tilde{U}(t_0, 0) = \exp[\Omega(t_0)] \equiv \exp[-it_0H_{\rm eff}], \tag{41}$$

where $H_{\rm eff} \equiv \frac{i}{t_0}\Omega(t_0)$, and $\Omega(t_0)$ can be computed using the (gate-appended) toggling-frame Hamiltonian $\tilde{H}(t)$ in Eq. (38). To bound the quantity $\|\tilde{U}(t_0, 0) - U_B(t_0)\|$, we make use of Lemma 3 in Appendix C, which gives:

$$\|\tilde{U}(t_0, 0) - U_B(t_0)\| \le t_0\|H_{\rm eff} - H_B\|. \tag{42}$$

Inserting the Magnus expansion $-it_0H_{\rm eff} = \Omega(t_0) = \sum_{n=1}^\infty \Omega_n(t_0)$ we find

$$\begin{aligned}\eta_{\rm DD} &\le t_0 \max_a \|H_{{\rm eff},a} - H_{B,a}\|\\ &= \max_a \|\sum_{n=1}^\infty \Omega_{n,a}(t_0) + it_0H_{B,a}\|\\ &\le \max_a \Big(\|\Omega'_{1,a}(t_0)\| + \sum_{n=2}^\infty \|\Omega_{n,a}(t_0)\|\Big),\end{aligned} \tag{43}$$

where $\Omega'_{1,a}(t) \equiv \Omega_{1,a}(t) + itH_{B,a}$. For a pulse sequence that achieves first-order decoupling with ideal zero-width pulses, $\Omega'_{1,a}(t_0)$ vanishes in the limit $\delta \to 0$. To derive a useful upper bound on the effective noise strength $\eta_{\rm DD}$, we will need good bounds on the other terms in Eq. (43).

C. The effective noise strength for a time-symmetric sequence

We say that the memory pulse sequence is time-symmetric (or "palindromic") if $\tilde{H}(t_{\rm DD}-t)=\tilde{H}(t)$ for $t\in[0,t_{\rm DD}]$. We will show in Appendix B that a time-symmetric pulse sequence that achieves first-order decoupling also achieves second-order decoupling. However, the time symmetry is broken if we construct the DD-protected gate by appending the gate pulse to the memory sequence, even if the memory sequence by itself is time-symmetric.

For the purpose of estimating the effective noise strength, we can nearly restore the time symmetry of the DD-protected gate by a simple trick (see Fig. 3). Recalling that our goal is to derive an upper bound on $\|\tilde{U}(t_0,0)-U_B(t_0)\|$, we observe that the unitary invariance of the operator norm implies

$$\|\tilde{U}(t_0,0)-U_B(t_0)\|=\|\tilde{U}(t_0,0)U_B^\dagger(\delta)-U_B(t_0-\delta)\|, \quad (44)$$

where δ is the width of the gate pulse, and $t_0=t_{\rm DD}+\delta$. Furthermore, we may regard $\tilde{U}(t_0,0)U_B^\dagger(\delta)$ as the time evolution operator generated by the Hamiltonian

$$H_M(t)\equiv\begin{cases}-H_B & t\in[0,\delta),\\ \tilde{H}(t-\delta) & t\in[\delta,T],\end{cases} \quad (45)$$

where $T=t_0+\delta$. If the memory sequence is time-symmetric, then $H_M(T-t)=H_M(t)$ for $t\in[\delta,T-\delta]$. Thus $H_M(t)$ is "nearly time-symmetric" in the interval $[0,T]$; the symmetry is broken only in the small intervals $[0,\delta]$ and $[T-\delta,T]$ at the beginning and end of $[0,T]$.

The unitary operator $\tilde{U}(t_0,0)U_B^\dagger(\delta)\equiv\exp[\Omega(T)]$ can be computed using the Magnus expansion for Hamiltonian $H_M(t)$. Viewing $\tilde{U}(t_0,0)U_B^\dagger(\delta)$ as being generated by the time-independent Hamiltonian $i\Omega(T)/(t_0-\delta)$ for time $t_0-\delta$, and again using Lemma 3 in Appendix C, we obtain instead of Eq. (43),

$$\begin{aligned}\eta_{\rm DD}&=\max_a\|\tilde{U}_a(t_0,0)U_{B,a}^\dagger(\delta)-U_{B,a}(t_0-\delta)\|\\ &\le\max_a\|\Omega_a(T)+i(T-2\delta)H_{B,a}\|\\ &=\max_a\Big(\|\Omega'_{1,a}(T)\|+\sum_{n=2}^\infty\|\Omega_{n,a}(T)\|\Big),\end{aligned} \quad (46)$$

where $\Omega'_{1,a}(T)$ is now defined as $\Omega'_{1,a}(T)\equiv\Omega_{1,a}(T)+i(T-2\delta)H_{B,a}$.

More generally, we say that the Hamiltonian $H_M(t)$ is nearly time-symmetric in $[0,T]$ if the time symmetry holds everywhere except in a small interval or the disjoint union of several small intervals. We denote by Δ the region in which the time symmetry is violated; thus

$$\begin{cases}H_M(T-t)=H_M(t) & \text{for } t\notin\Delta\\ H_M(T-t)\neq H_M(t) & \text{for } t\in\Delta.\end{cases} \quad (47)$$

We also use the same symbol $\Delta(\ll T)$ to denote the total length of this region. Thus $\Delta=0$ for a perfectly time-symmetric sequence. In what follows, we will sometimes say that the pulse sequence realizing a DD-protected gate is "time-symmetric" if the corresponding memory sequence is time-symmetric, even though the time symmetry may be broken by the gate pulse appended to the memory sequence. We say that the memory sequence and the DD-protected gates are "general" if the memory sequence has no special time symmetry properties. We note, though, that even in the general case, we may say that a pulse sequence with N pulses, each of duration δ, is nearly time-symmetric with $\Delta=2N\delta$, since of course $H_M(T-t)=H_M(t)$, if the control Hamiltonian vanishes at time t and at time $T-t$.

V. EFFECTIVE NOISE STRENGTH AND THRESHOLD CONDITIONS

In this section, we state some of our conclusions concerning the effective noise strength $\eta_{\rm DD}$ achieved by dynamical decoupling, and the implications for fault-tolerant quantum computing. Derivations will be postponed until Sec. VII. Here we focus on results derived using the Magnus expansion; results relating $\eta_{\rm DD}$ to properties of bath correlation functions are discussed in Sec. X.

A. Bounds on Magnus expansion

Combining Eq. (43) and Eq. (46), we can state our upper bound on the effective noise strength $\eta_{\rm DD}$ as

$$\eta_{\rm DD}\le\|\Omega'_1(T)\|+\sum_{n=2}^\infty\|\Omega_n(T)\|, \quad (48)$$

where $\Omega'_1(T)\equiv\Omega_1(T)+i(T-2\Gamma)H_B$, $T\equiv t_0+\Gamma$, and the maximum over all locations is understood. The Magnus expansion $\Omega(T)=\sum_{n=1}^\infty\Omega_n(T)$ is computed using the Hamiltonian

$$H_M(t)\equiv\begin{cases}-H_B & t\in[0,\Gamma),\\ \tilde{H}(t-\Gamma)=H_B+\tilde{H}_{\rm err}(t-\Gamma) & t\in[\Gamma,T].\end{cases} \quad (49)$$

For the general case, in which we are not trying to exploit the time symmetry of the memory sequence, we choose $\Gamma=0$. For the nearly-time-symmetric case we choose $\Gamma=\delta$, and H_M is time-symmetric in the interval $[\delta,T-\delta]$.

If the memory sequence achieves first-order decoupling in the limit $\delta\to0$, then $\|\Omega'_1(T)\|$ vanishes apart from finite-width corrections. The nth-order Magnus term $\Omega_n(T)$ for $n\ge2$ satisfies $\|\Omega_n(T)\|=O\left((\epsilon T)^n\right)$, because $\|H_M(t)\|\le\epsilon$, and the integral $\Omega_n(T)$ can be bounded by the volume of the integration region times an upper bound on the integrand. In fact, this estimate can be improved to $\|\Omega_n(T)\|=O\left((JT)(\epsilon T)^{n-1}\right)$, because $H_M(t)$ has the form $\pm H_B+H'(t)$ where $H'(t)$ is either 0 or $\tilde{H}_{\rm err}(t)$; therefore $\|[H_M(t_1),H_M(t_2)]\|=O(J\epsilon)$, since $\|H'(t)\|\le J$ and H_B commutes with itself.

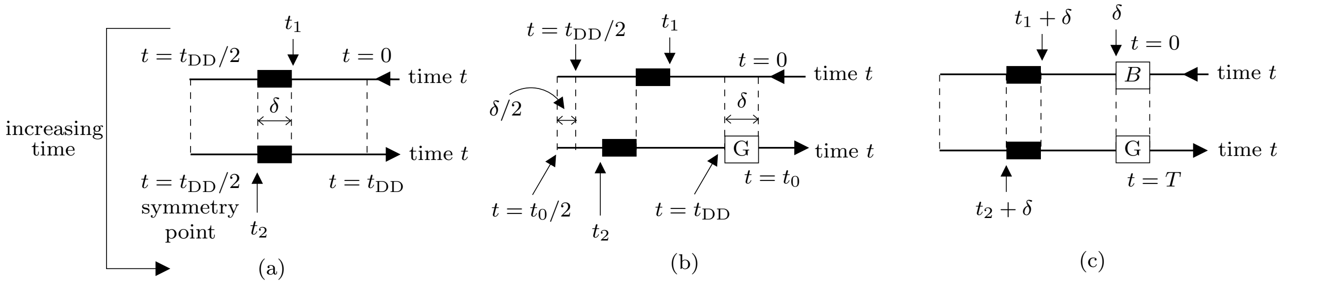

FIG. 3: Schematic representation of $H_M(t)$ for time-symmetric DD pulse sequences. The time axis is bent in half, with time flowing counterclockwise from the upper right to the lower left, so that times aligned on the upper and lower branches are related by time symmetry. (a) Two pulses (marked as black boxes) in a time-symmetric memory sequence with $\tilde{H}(t_{\rm DD} - t) = \tilde{H}(t)$; the pulse on the bottom branch is the time-reversed version of the pulse on the top branch. (b) Appending the gate pulse (box G) to the memory sequence spoils the time symmetry; the black pulses on the upper and lower branches are no longer aligned. (c) Appending fictitious time evolution governed by $-H_B$ during $t \in [0, \delta]$ (box B) restores the time symmetry of the memory sequence for $t \in [\delta, T - \delta]$, where $T = t_0 + \delta$.

We anticipate, then, that at any location a, the terms in the Magnus expansion can be bounded as

$$\|\Omega_1'(T)\| \le C_1(JT); \quad (50a)$$

$$\|\Omega_n(T)\| \le C_n(JT)(\epsilon T)^{n-1}, \quad n = 2, 3, 4; \quad (50b)$$

$$\sum_{n=5}^{\infty} \|\Omega_n(T)\| \le C_5(JT)(\epsilon T)^4, \quad (50c)$$

where C_1, C_2, C_3, C_4, C_5 are constants. Note that the last of these results bounds the sum of all high-order Magnus terms with $n \ge 5$. Combining Eq. (48) and Eq. (50) we find

$$\eta_{\rm DD} \le (JT) \sum_{n=1}^{5} C_n (\epsilon T)^{n-1}. \quad (51)$$

The constants C_n, derived in Sec. VII, are listed in Table I for both general and time-symmetric pulse sequences. Our value of C_5, obtained by bounding an infinite series, holds only for $\epsilon T \le 0.54$, a condition likely to be satisfied when DD works effectively. If desired, tighter bounds can be derived on the nth order terms with $n \ge 5$ using results from Sec. VII. However, we judge Eq. (50c) to be good enough for our purposes, since this bound on the sum of higher-order terms is already rather small for $\epsilon T \ll 1$, as in typical cases of interest.

Our bound on $\Omega_1(T)$ is not improved by invoking time symmetry, but for $n = 2, 3, 4$, the bounds listed in Table I are tighter for the time-symmetric case than the general case, assuming $\Delta/T \ll 1$. In fact, C_2 and C_4 vanish in the limit $\Delta/T \to 0$, reflecting the property that all even-order terms in the Magnus expansion vanish when the time symmetry is exact. For the time-symmetric case, we derive bounds on C_n for even $n \ge 6$ in Appendix D, but these results were not used in our estimate of C_5.

	General	**Nearly time-symmetric**
C_1	$2N\delta/T$ in general, $N\delta/T$ if pulses are regularly spaced in time,	
C_2	1	$4\left(\frac{\Delta}{T}\right)\left(1 - \frac{1}{2}\frac{\Delta}{T}\right)$
C_3	4/9	$(2/9)\left\{1 + 15\left(\frac{\Delta}{T}\right)\left(1 - \frac{14}{15}\frac{\Delta}{T}\right)\right\}$
C_4	11/9	$56\left(\frac{\Delta}{T}\right)\left(1 - \frac{1}{2}\frac{\Delta}{T}\right)$
C_5	9.43	

TABLE I: Constants $\{C_n\}$ appearing in the upper bound Eq. (50) on the Magnus expansion, for the general case and the nearly-time-symmetric case. N denotes the total number of pulses in the DD-protected gate, δ is the pulse width, $T = t_0$ in the general case, and $T = t_0 + \delta$ in the nearly-time-symmetric case. For the nearly-time-symmetric case, Δ is the size of the small region in which the time symmetry is violated.

B. Noise suppression threshold

DD-protected gates will outperform unprotected gates if the noise is weak enough. In a circuit of unprotected gates, each gate is realized by a single pulse, where the pulses are separated in time by the pulse interval τ_0. For the noise model of Sec. II, the effective noise strength for this computation may be expressed as [30, 31]

$$\eta = \left(\max_a \|H_{SB,a}\|\right) \tau_0. \quad (52)$$

Eq. (52) is not derived using the Magnus expansion; rather it follows directly from Lemma 3 in Appendix C. If we assume that $H_S^0 = 0$, Eq. (52) becomes

$$\eta = J\tau_0. \quad (53)$$

We say that the noise model satisfies the *noise suppression threshold condition* if the effective noise strength can be reduced by using DD-protected gates instead, *i.e.*, if

$$\eta_{\rm DD} < \eta. \quad (54)$$

In our noise model, this condition can be expressed in terms of the parameters $\epsilon\tau_0$, δ/τ_0 and τ_0/t_0.

For example, continuing to assume that $H_S^0 = 0$, suppose in addition that δ/τ_0 is negligible and ϵT is small enough so that the Magnus expansion is well-approximated by the lowest-order nonzero term. Then, in the general (non-time-symmetric) case, using $C_1 = 0$ and $C_2 = 1$, we can approximate $\eta_{\rm DD}$ by

$$\eta_{\rm DD} \simeq (JT)(\epsilon T) = \left(\frac{J\tau_0}{\tau_0/t_0}\right)\left(\frac{\epsilon\tau_0}{\tau_0/t_0}\right); \tag{55}$$

we use the $\simeq$ symbol to emphasize that higher order corrections in δ/τ_0 and ϵt_0 are neglected. Note that $\eta_{\rm DD}$ depends on the norm of the bath Hamiltonian $\|H_B\|$ (which contributes to ϵ), while η does not. The noise suppression threshold condition $\eta_{\rm DD} < \eta = J\tau_0$ is satisfied for

$$\epsilon\tau_0 \lesssim \left(\frac{\tau_0}{t_0}\right)^2, \tag{56}$$

or

$$\epsilon\tau_0 \lesssim N^{-2} \tag{57}$$

for a sequence of N equally spaced pulses. As the pulse sequence grows, the duration t_0 of DD-protected gates increases relative to the duration τ_0 of unprotected gates, and Eq. (56) imposes a stronger restriction on ϵ.

In the limit of zero-width pulses, a time-symmetric pulse sequence that achieves first-order decoupling achieves second-order decoupling as well, so that $C_1 = C_2 = 0$. Imposing time symmetry may lengthen the pulse sequence; we denote the duration of a DD-protected time-symmetric gate by t_0', to contrast with the duration t_0 of the gate when the pulse sequence is not time-symmetric. In the time-symmetric case, the effective noise strength becomes (assuming $\delta = 0$ and thus $T = t_0$, and using $C_3 = 2/9$)

$$\eta_{\rm DD} \simeq \frac{2}{9}(JT)(\epsilon T)^2 = \frac{2}{9}\left(\frac{J\tau_0}{\tau_0/t_0'}\right)\left(\frac{\epsilon\tau_0}{\tau_0/t_0'}\right)^2. \tag{58}$$

Therefore the noise suppression threshold condition $\eta_{\rm DD} < J\tau_0$ is satisfied if

$$\epsilon\tau_0 \lesssim \frac{3}{\sqrt{2}}\left(\frac{\tau_0}{t_0'}\right)^{3/2}, \tag{59}$$

or

$$\epsilon\tau_0 \lesssim \frac{3}{\sqrt{2}}\left(N'\right)^{-3/2} \tag{60}$$

for the case of N' equally spaced pulses. Even though $t_0' > t_0$, Eq. (59) places a less stringent condition on ϵ than Eq. (56), provided $t_0'/t_0 \lesssim (9t_0/2\tau_0)^{1/3}$. We emphasize again that Eq. (56) and Eq. (59) are derived using lowest-order approximations in an expansion in δ/τ_0 and ϵ.

The expression Eq. (55) for $\eta_{\rm DD}$ indicates that to achieve effective noise suppression we should favor short DD pulse sequences (with t_0/τ_0 not too large) to minimize the exposure to noise during the DD-protected gate. On the other hand Eq. (58) illustrates that a longer pulse sequence can pay off if it allows us to achieve higher-order decoupling. These results exemplify a more general tradeoff between shorter sequences and better decoupling that must be optimized to design DD-protected gates with the smallest possible effective noise strength. The tradeoff is also manifested by the analysis in Sec. VIII of concatenated DD pulse sequences.

C. Accuracy threshold and overhead cost

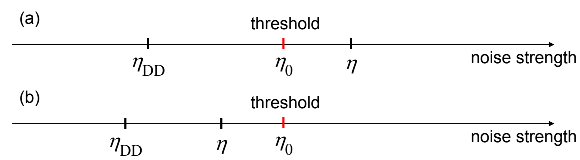

FIG. 4: Two scenarios where DD-protected gates outperform unprotected gates. (a) Quantum computing is scalable with DD-protected gates, but not with unprotected gates. (b) Quantum computing is scalable with either DD-protected gates or with unprotected gates, but DD reduces the overhead cost of fault tolerance.

A quantum computation unprotected by DD is scalable if the noise strength of unprotected gates is below the accuracy threshold, $\eta < \eta_0$. For DD-protected gates, the accuracy threshold condition becomes $\eta_{\rm DD} < \eta_0$. If the noise suppression threshold condition is satisfied, so that $\eta_{\rm DD} < \eta$, it may be that $\eta > \eta_0$ and $\eta_{\rm DD} < \eta_0$; in that case, arbitrarily large quantum circuits can be simulated accurately with DD-protected gates, but not with unprotected gates. This is illustrated in Fig. 4(a).

Even when $\eta < \eta_0$, DD may reduce the overhead cost of fault-tolerant quantum computing if $\eta_{\rm DD} < \eta$ (Fig. 4(b)). Suppose that we wish to simulate an ideal quantum circuit containing L gates. If our noisy gates have noise strength $\bar{\eta}$, which is below the threshold value η_0, the simulation is possible using L^* noisy gates where [31]

$$\frac{L^*}{L} = \left(\frac{\log(c\eta_0 L/\theta)}{\log(\eta_0/\bar{\eta})}\right)^a; \tag{61}$$

here c and a are constants, and θ is the "error" in the simulation (the L^1 distance between the ideal probability distribution of outcomes and the simulated distribution). Denote by $L^*_{\rm un}$ the number of pulses in the fault-tolerant circuit built from unprotected gates, and by $L^*_{\rm DD}$ the number of pulses in the fault-tolerant circuit built from DD-protected gates, and suppose that each

DD-protected gate uses N pulses, while each unprotected gate uses a single pulse. Then the ratio

$$\frac{L^*_{\rm DD}}{L^*_{\rm un}} = N\left(\frac{\log(\eta_0/\eta)}{\log(\eta_0/\eta_{\rm DD})}\right)^a \tag{62}$$

is independent of the size L of the simulated circuit. If using DD substantially improves the effective noise strength, $L^*_{\rm DD}/L^*_{\rm un}$ may be small, especially if η is only slightly below the threshold value η_0. Even though each DD-protected gates requires multiple pulses, the total number of pulses used in the simulation may be reduced, because DD improves the gate accuracy.

Of course, we have reached this conclusion using the local-bath assumption, which allows us to assign a well-defined effective noise strength to the DD-protected gate. Furthermore our results are useful only if the Hamiltonian of the local bath has finite norm (so that $\epsilon < \infty$). However, we will see that the correlation function analysis in Sec. X can provide useful upper bounds on $\eta_{\rm DD}$ even if ϵ is infinite.

D. Slow preparations and measurements

Another drawback of this analysis is that our model of qubit preparations and measurements may be unrealistic in some physical situations. In our estimates of the effective noise strength in a DD-protected quantum computation, we have treated preparations and measurements like gates. We have assumed that each preparation and measurement location in the circuit, like each gate location, has duration t_0. A DD-protected preparation location consists of a preparation taking time δ followed by a DD memory sequence, and a DD-protected measurement location consists of a DD memory sequence followed by a measurement taking time δ. Thus we have assumed that the preparations and measurements are just as fast as the pulses. In some physical systems, however, preparations and measurements are relatively slow; in solid-state devices, for example, the measurement time can be orders of magnitude longer than the gate time.

If the actual time $\bar{\delta}$ required for a preparation or measurement is longer than the pulse width δ but still short compared to the pulse interval τ_0, then we could still try to improve measurements and preparations using DD sequences. If it makes sense to model the noise during a preparation or measurement as we have modeled the noise in the pulses, then we could modify our analysis by using the measurement width $\bar{\delta}$ in estimating the effective noise strength $\eta_{\rm DD}$ at preparation and measurement locations, while continuing to use the pulse width δ in estimating $\eta_{\rm DD}$ at gate locations. But if $\bar{\delta} \gtrsim \tau_0$, or more generally if the noise in preparations and measurements is modeled much differently than the noise in gates, then it may be more appropriate to consider the preparation/measurement noise strength to be a separate parameter in the analysis, not necessarily related to the parameters J and ϵ that characterize the noise Hamiltonian described in Sec. II and appear in the estimate of $\eta_{\rm DD}$ at gate locations.

Measurement locations might be much noisier than gate locations because gates can be improved using sequences of fast DD pulses, while slow measurements cannot be improved by DD. Or measurements might be noisier than gates for other quite different reasons. Previous work has shown that scalable fault-tolerant quantum computing is still possible, and that the accuracy threshold is not much affected, when measurements are much *slower* than gates [48]. What deserves further study, though, is how fault-tolerant circuit design can be optimized when measurements are much *noisier* than gates.

VI. EXAMPLES

Now we will analyze the effectiveness of several different DD pulse sequences, applying the results from Sec. V A. We adopt a noise model that includes only single-qubit errors acting on the system; thus the noise Hamiltonian is

$$H = H_B + \sum_{i,\alpha} \sigma^{(i)}_\alpha \otimes B^{(i)}_\alpha, \tag{63}$$

where i labels the qubits, $\sigma^{(i)}_\alpha$ for $\alpha = x, y, z$ are the Pauli operators acting on qubit i, and

$$H_B = \mathbb{I}_S \otimes B_0. \tag{64}$$

In some realistic situations, such as electron-spin qubits interacting with a nuclear spin bath [17, 18, 19, 40], such single-qubit errors are the dominant noise in the system.

In principle, $H_{\rm err}$ could also contain errors that act collectively on several qubits at once; for example, errors acting jointly on two qubits might be expected to occur during the execution of a two-qubit gate. Efficient DD pulse sequences can be constructed that suppress multi-qubit errors [49, 50], but in this Section we limit our attention to single-qubit noise and pulse sequences that combat it. The more general results in Sec. V A can also be applied to other models that include multi-qubit noise and to the corresponding pulse sequences that achieve first-order decoupling for such noise.

We will discuss three different DD pulse sequences that can suppress the single-qubit noise. The first is the simplest DD scheme that protects against arbitrary single-qubit errors. The second is a time-symmetric sequence that achieves second-order decoupling in the limit of zero-width pulses. The third is the Eulerian DD scheme, which is more robust against pulse errors than the other schemes.

A. Universal decoupling sequence

The shortest pulse sequence that suppresses arbitrary single-qubit errors is called the "universal decoupling sequence" [6, 10], or "XY-4" in the NMR literature [51]. For this sequence, the unitary operator generated by the control Hamiltonian, acting on a single qubit, can be expressed as

$$U_c(t_{\rm DD}) = Z\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}. \tag{65}$$

The notation in Eq. (65) is meant to convey that one complete cycle of the memory sequence contains four equally spaced pulses (each of width δ) that successively apply the Pauli operators X, Z, X, Z, where $X = \sigma_x$ and $Z = \sigma_z$; therefore the product of the four Pauli operators is $-\mathbb{I}$. Each $\mathbb{I}$ in Eq. (65) represents trivial evolution during the pulse interval of width $\tau_0 - \delta$. The total duration of the pulse sequence is $t_{\rm DD} = 4\tau_0$.

This sequence achieves first-order decoupling. In the limit of zero-width pulses, the toggling frame Hamiltonian is

$$\tilde{H}(t) = U_c^\dagger(t) H U_c(t) = \begin{cases} \mathbb{I}H\mathbb{I} \;\; = H_B + X\otimes B_X + Y\otimes B_Y + Z\otimes B_Z & \text{for } t\in[0,\tau_0), \\ XHX = H_B + X\otimes B_X - Y\otimes B_Y - Z\otimes B_Z & \text{for } t\in[\tau_0,2\tau_0), \\ YHY = H_B - X\otimes B_X + Y\otimes B_Y - Z\otimes B_Z & \text{for } t\in[2\tau_0,3\tau_0), \\ ZHZ = H_B - X\otimes B_X - Y\otimes B_Y + Z\otimes B_Z & \text{for } t\in[3\tau_0,4\tau_0), \end{cases} \tag{66}$$

and we find

$$\begin{aligned}\Omega_1(t_{\rm DD}) &= -i\int_0^{t_{\rm DD}} dt \tilde{H}(t) \\ &= -i\tau_0\left(\mathbb{I}H\mathbb{I} + XHX + YHY + ZHZ\right) \\ &= -it_{\rm DD}H_B, \end{aligned} \tag{67}$$

a pure-bath term. The first-order Magnus term (up to the factor $-it_{\rm DD}$) is the Pauli-group average of the noise Hamiltonian H, which commutes with any nontrivial Pauli operator acting on the system qubit.

In a DD-protected gate, the final pulse in the universal decoupling sequence is modified by combining with the gate pulse. For a single-qubit gate, the pulse sequence realizing the gate G is

$$U_c(t_0) = (GZ)\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}, \tag{68}$$

where now GZ represents a single pulse with duration δ and $t_0 = t_{\rm DD}$. In a two-qubit gate, the universal pulse sequence is applied in parallel to both qubits, except that the final pulse $Z\otimes Z$ in the memory sequence is replaced by the two-qubit pulse $G(Z\otimes Z)$.

To estimate the effective noise strength $\eta_{\rm DD}$, we note that the total number of pulses is $N = 4$, and that $\tau_0/t_0 = 1/4$. From the bounds in Eq. (50) and Table I (for the case where the sequence is not time symmetric) we obtain

$$\eta_{\rm DD} = (4J\tau_0)\left[\frac{\delta}{\tau_0} + 4\epsilon\tau_0 + \frac{4}{9}(4\epsilon\tau_0)^2 + \frac{11}{9}(4\epsilon\tau_0)^3 + 9.43(4\epsilon\tau_0)^4\right], \tag{69}$$

where we have used $C_1 = N\delta/t_0$ because the pulses are regularly spaced in time. Note that the parameters J and ϵ include sums over all qubits in the set $\mathcal{Q}_a$ that participate in the gate at location a in the circuit.

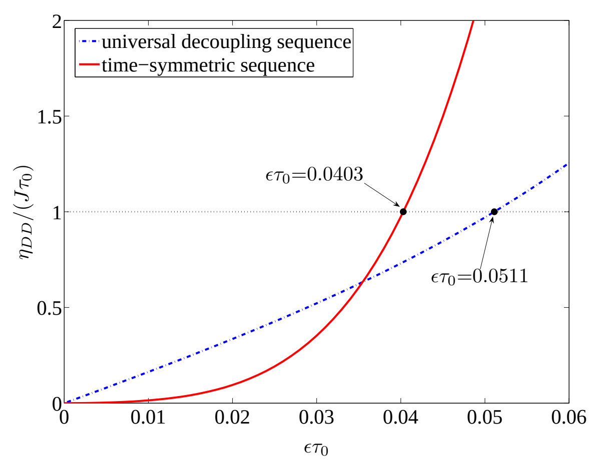

FIG. 5: Plot of $\eta_{\rm DD}/\eta$ versus $\epsilon\tau_0$ for the universal decoupling sequence (Eq. 69) and for the time-symmetric sequence (Eq. 74), assuming zero-width pulses. The noise strength for the DD-protected gate is weaker than the noise strength for the unprotected gate for $\epsilon\tau_0 < 0.0511$ in the case of the universal decoupling sequence, and for $\epsilon\tau_0 < 0.0403$ in the case of the time-symmetric sequence. For $\epsilon\tau_0$ sufficiently small, using the time-symmetric sequence reduces the noise strength further than the universal decoupling sequence.

In Fig. 5, $\eta_{\rm DD}/\eta$ (where $\eta = J\tau_0$) is plotted as a function of $\epsilon\tau_0$, in the limit $\delta/\tau_0 \to 0$. The noise suppression threshold condition $\eta_{\rm DD} < \eta$ is satisfied when

$$\epsilon\tau_0 < 0.0511. \tag{70}$$

In the limit $\epsilon\tau_0 \to 0$, the noise suppression threshold condition is satisfied for

$$\frac{\delta}{\tau_0} < \frac{1}{4}. \tag{71}$$

B. Time-symmetric sequence

We can construct a time-symmetric DD sequence by composing two copies of the universal decoupling sequence — first we perform the sequence in the forward direction, and then run it backwards in time. For zero-width pulses, using the same notation as in Eq. (65), in which $\mathbb{I}$ represents trivial evolution for time τ_0 between successive pulses, this sequence can be expressed as

$$U_c(t_{\rm DD}) = \mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}, \tag{72}$$

where we have combined the two Z operators in the middle into a zero-width identity "pulse" (not shown in Eq. (72)). The total duration of the pulse sequence is $t_{\rm DD} = 8\tau_0$, twice as long as the universal decoupling sequence. Like the universal decoupling sequence, this sequence achieves first-order decoupling. In addition, it satisfies the time-symmetry property $U_c(t_{\rm DD}-t) = U_c(t)$, so that the toggling-frame Hamiltonian obeys $\tilde{H}(t_{\rm DD} - t) = \tilde{H}(t)$, and thus this pulse sequence achieves second-order decoupling as well. This pulse sequence is known in the NMR literature as "XY-8" [52].

For finite-width pulses, we modify our notation to emphasize that the second half of the sequence is the time reverse of the first half. We write

$$U_c(t_{\rm DD}) = \mathbb{I}X^{(-)}\mathbb{I}Z^{(-)}\mathbb{I}X^{(-)}\mathbb{I}\mathbb{I}_\delta\mathbb{I}X^{(+)}\mathbb{I}Z^{(+)}\mathbb{I}X^{(+)}\mathbb{I}. \tag{73}$$

Now, each $\mathbb{I}$ represents trivial evolution for time $(\tau_0 - \delta)$. The $\mathbb{I}_\delta$ operator in the middle represents trivial evolution for time δ, arising from combining two Z pulses. It might seem more natural to use $\mathbb{I}_{2\delta}$ instead, matching the total duration of two Z pulses each with width δ, but we choose the sequence Eq. (73) so that our upper bound on $\Omega_3(T)$, the dominant Magnus term when δ/τ_0 is negligible, will have a simple form. Since $\delta/t_{\rm DD}$ is small anyway, it does not matter much which of these sequences we choose. $X^{(\pm)}$ and $Z^{(\pm)}$ represent finite-width pulses implementing X and Z. Before the midpoint of the sequence at $t = t_{\rm DD}/2$, the X pulses are executed using the constant Hamiltonian H_{P_X} such that $X = \exp(-i\delta H_{P_X})$ and the Z pulse is executed using H_{P_Z} such that $Z = \exp(-i\delta H_{P_Z})$, assuming the pulses are perfectly rectangular. After the midpoint, the universal decoupling sequence runs backwards; X is executed using $-H_{P_X}$ and Z using $-H_{P_Z}$. Thus, $U_c(t_{\rm DD} - t) = U_c(t)$.

Appending the gate pulse to this memory sequence breaks the time symmetry, which can be nearly restored using the trick explained in Sec. IV C. The region Δ in which the time symmetry is violated is the union of two intervals: the duration of the gate pulse, and its image under time reversal, during which evolution is governed by the Hamiltonian $-H_B$. Thus $\Delta = 2\delta$ (recall that we use Δ to denote both the region and its size). The DD-protected gate contains $N = 8$ pulses (seven pulses in the memory sequence, including the identity pulse in the middle, plus the gate pulse) and has duration $t_0 = 8\tau_0$, so that $\tau_0/t_0 = 1/8$ and $T = t_0 + \delta$. From the bounds in Eq. (50) and Table I (for the case where the sequence is nearly time-symmetric) we obtain an estimate of the effective noise strength $\eta_{\rm DD}$ of the DD-protected gates; we may use $C_1 = N\delta/T \le \delta/\tau_0$ because the pulses are regularly spaced in time.

In the limit of zero-width pulses ($\delta/\tau_0 \to 0$), the effective noise strength becomes

$$\eta_{\rm DD} = (8J\tau_0)\left[\frac{2}{9}(8\epsilon\tau_0)^2 + 9.43(8\epsilon\tau_0)^4\right]; \tag{74}$$

$\eta_{\rm DD}/\eta$ is plotted in Fig. 5. The noise suppression threshold condition $\eta_{\rm DD} < \eta$ is satisfied when

$$\epsilon\tau_0 < 0.0403 \tag{75}$$

This condition is more stringent than for the universal decoupling sequence, which is not surprising since the time-symmetric sequence is twice as long. As Fig. 5 illustrates, the time-symmetric sequence becomes more advantageous when $\epsilon\tau_0$ is small, as is likely to be the case when $\eta_{\rm DD}$ is below the accuracy threshold η_0. In the limit $\epsilon\tau_0 \to 0$, only C_1 survives, and we find $\eta_{\rm DD} \le 8\eta(\delta/\tau_0)$; thus the noise suppression threshold condition is satisfied for

$$\frac{\delta}{\tau_0} < \frac{1}{8}. \tag{76}$$

The largest permissible pulse width is half as large as in the case of the universal decoupling sequence (eq. (71)) because the time-symmetric sequence is twice as long.

Using Eq. (62) and the expressions for $\eta_{\rm DD}$ in Eq. (69) (with $\delta/\tau_0 = 0$) and Eq. (74), we plot in Fig. 6 the ratio $L^*_{\rm DD}/L^*_{\rm un}$ versus $\epsilon\tau_0$ for both the universal decoupling sequence and the time-symmetric sequence. Here, just to illustrate the idea that DD can drastically reduce the overhead requirements for fault-tolerant quantum computing, we have assumed $\eta_0/\eta = 2$, and we have taken the value $a = \log_2(291) \approx 8.18$ from [31] (291 is the number of locations, including measurements and preparations, contained in the fault-tolerant CNOT gadget constructed in [31]). Because the noise strength for the unprotected gate is close to the threshold value, and because $\epsilon\tau_0$ is well below the noise suppression threshold for each DD sequence in the range plotted, the reduction in the number of pulses achieved by using DD-protected gates is substantial. Furthermore, although the time-symmetric sequence is longer than the universal decoupling sequence, the time-symmetric sequence reduces the total number of pulses more effectively than the universal sequence, by more than two orders of magnitude for $\epsilon\tau_0 < 10^{-2}$.

For some noise models, the value of $\eta_{\rm DD}$ derived by our general arguments may be overly pessimistic. For example, using the time-symmetric sequence Eq. (72), we computed $\Omega_3(T)$ for a single-qubit system coupled to an n-spin bath in an external magnetic field, assuming an isotropic Heisenberg interaction between the system qubit and each bath spin. The ratio of the bound from

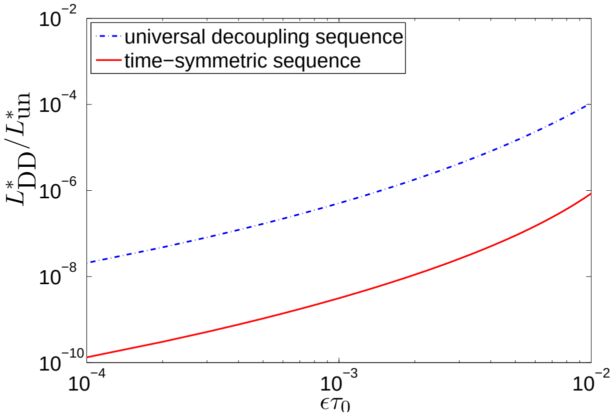

FIG. 6: Plot of $L^*_{\mathrm{DD}}/L^*_{\mathrm{un}}$ (Eq. (62)) versus $\epsilon\tau_0$ for the universal decoupling sequence and the time-symmetric sequence, in the limit $\delta/\tau_0 \to 0$. We have assumed $\eta_0/\eta = 2$, and have used the value $a = \log_2(291) \approx 8.18$ appropriate for the fault-tolerant gadget constructions in [31].

Eq. (50) and Table I to the actual value of $\|\Omega_3(T)\|$ for this model is plotted in Fig. 7 as a function of $\beta\tau_0$, for $H_S^0 = 0$ and $\delta = 0$. The bound is larger than the actual value by at least a factor of 20.

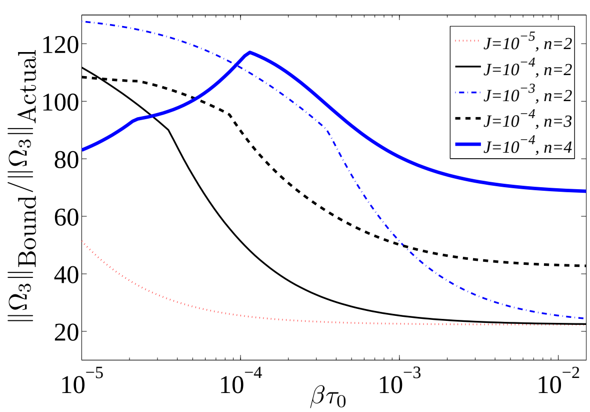

FIG. 7: Plot of $\|\Omega_3\|_{\mathrm{Bound}}/\|\Omega_3\|_{\mathrm{Actual}}$ versus $\beta\tau_0$ for the time-symmetric DD sequence Eq. (72). The noise Hamiltonian is $H = H_B + H_{SB}$ ($H_S^0 = 0$), where $H_B = (\beta/2)\sum_i \sigma_i^z$ and $H_{SB} = (J/4)\sum_{\alpha=x,y,z}\sigma_S^\alpha \otimes \left(\sum_{i=1}^n \sigma_i^\alpha\right)$; the index i labels the bath spins. Here $\|\Omega_3\|_{\mathrm{Bound}}$ is computed using Eq. (50) and Table I (where $\delta = 0$ and $T = 8\tau_0$), while $\|\Omega_3\|_{\mathrm{Actual}}$ is computed by evaluating the integral in Eq. (28) exactly. The kinks in the plots arise because the operator norm can have a discontinuous first derivative when eigenvalues cross.

C. Eulerian decoupling sequences

The effects of finite pulse width and other pulse imperfections can be suppressed by using an "Eulerian" memory sequence [7]. In Eulerian decoupling, the operator applied by each pulse is the generator of a finite group, and $U_c(t)$ traverses an Euler cycle in the Cayley graph of this group. As a result, the error Hamiltonian is group averaged and first-order decoupling is maintained even when the pulses have (reproducible) imperfections. We will describe a simple Eulerian memory sequence here; see [7] for a more general discussion.

A simple Eulerian memory sequence protecting against single-qubit noise is [7]

$$U_c(t_{\mathrm{DD}}) = X\mathbb{I}Z\mathbb{I}X\mathbb{I}Z\mathbb{I}Z\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}. \tag{77}$$

Here the pulses are equally spaced in time; each $\mathbb{I}$ operators represents the same time interval, and the spacing between the start of two consecutive pulses is τ_0. This sequence looks like two repetitions of the universal decoupling sequence, except that the X and Z pulses are swapped in the second repetition. In contrast to the time-symmetric sequence Eq. (73), we use the same Hamiltonian H_{P_X} to execute each X pulse, rather than reversing the sign of the Hamiltonian during the second half of the sequence; similarly we use the same Hamiltonian H_{P_Z} to execute each Z pulse.

Without making any assumption about the pulse widths or shapes (except for assuming that all X pulses are alike and that all Z pulses are alike), we may express the unitary evolution operator over the pulse interval of duration τ_0 as $u_X(t)$ for an X pulse and $u_Z(t)$ for a Z pulse. Then, for $t \in [0, t_{\mathrm{DD}} \equiv 8\tau_0]$, $U_c(t)$ becomes

$$U_c(t) = \begin{cases} u_X(t)\mathbb{I} & t \in [0,\tau_0) \\ u_Z(t-\tau_0)X & t \in [\tau_0, 2\tau_0) \\ u_X(t-2\tau_0)(iY) & t \in [2\tau_0, 3\tau_0) \\ u_Z(t-3\tau_0)(-Z) & t \in [3\tau_0, 4\tau_0) \\ u_Z(t-4\tau_0)(-\mathbb{I}) & t \in [4\tau_0, 5\tau_0) \\ u_X(t-5\tau_0)(-Z) & t \in [5\tau_0, 6\tau_0) \\ u_Z(t-6\tau_0)(iY) & t \in [6\tau_0, 7\tau_0) \\ u_X(t-7\tau_0)X & t \in [7\tau_0, 8\tau_0) \end{cases}. \tag{78}$$

The first-order Magnus term $\Omega_1(t_{\mathrm{DD}})$ can be expressed in terms of effective Hamiltonians H_X and H_Z, obtained by averaging the Hamiltonian over an X or Z pulse respectively:

$$\tau_0 H_X \equiv \int_0^{\tau_0} dt\; u_X^\dagger(t) H u_X(t),$$
$$\tau_0 H_Z \equiv \int_0^{\tau_0} dt\; u_Z^\dagger(t) H u_Z(t). \tag{79}$$

Since u_X and u_Z act only on the system, they commute with the bath Hamiltonian H_B; while averaging over the pulse alters H_{err}, it has no effect on H_B. Therefore we find that

$$\begin{aligned}\Omega_1(t_{\rm DD}) &= \int_0^{t_{\rm DD}} dt\ \tilde{H}(t) = \int_0^{t_{\rm DD}} dt\ U_c^\dagger(t) H U_c(t)\\ &= \tau_0\left(H_X + XH_XX + YH_XY + ZH_XZ\right)\\ &+ \tau_0\left(H_Z + XH_ZX + YH_ZY + ZH_ZZ\right)\\ &= 8H_B\tau_0;\end{aligned} \tag{80}$$

thus $\Omega_1(t_{\rm DD})$ is a pure bath term. To derive the last line of Eq. (80), we have used the property $H + XHX + YHY + ZHZ = 4H_B$ (as in Eq. (67)). We conclude that first-order decoupling is perfectly attained irrespective of the pulse shape, as long as the *same* $u_{X(Z)}(t)$ is applied for every $X(Z)$ pulse, and the *integrated* pulses are exactly right.

Adding a gate pulse G, by combining G with the final X pulse of the Eulerian memory sequence, introduces an error depending on the width of the final pulse. However, because this nonvanishing contribution to $\Omega_1(T)$ arises only from the final pulse, it does not depend on the length of the memory sequence. Other contributions to $\Omega(T)$ that depend on pulse shapes, in the second order of the Magnus expansion and beyond, are suppressed by additional factors of $\epsilon\tau_0$.

The contributions that depend on the pulse shape can be further suppressed by making the Eulerian memory sequence time-symmetric. Consider, for example, the sequence

$$\begin{aligned}&U_c(t_{\rm DD})\\ &= X^{(-)}\mathbb{I}Z^{(-)}\mathbb{I}X^{(-)}\mathbb{I}Z^{(-)}\mathbb{I}Z^{(-)}\mathbb{I}X^{(-)}\mathbb{I}Z^{(-)}\mathbb{I}X^{(-)}\mathbb{I}\\ &\times \mathbb{I}X^{(+)}\mathbb{I}Z^{(+)}\mathbb{I}X^{(+)}\mathbb{I}Z^{(+)}\mathbb{I}Z^{(+)}\mathbb{I}X^{(+)}\mathbb{I}Z^{(+)}\mathbb{I}X^{(+)},\end{aligned} \tag{81}$$

where the control Hamiltonian is chosen such that $u_{X^{(-)}}(t_{\rm DD}-t) = u_{X^{(+)}}(t)$ and $u_{Z^{(-)}}(t_{\rm DD}-t) = u_{Z^{(+)}}(t)$. Because this sequence obeys the time symmetry condition $U_c(t_{\rm DD}-t) = U_c(t)$, the even-order Magnus terms vanish. Furthermore, because Eq. (81) is just two copies of the Eulerian sequence Eq. (77), the first running backward in time and the second running forward, the sequence achieves first-order decoupling for any pulse shape. Corrections depending on the pulse shape enter only in third order and beyond. Of course, making the Eulerian sequence time-symmetric (or making the time-symmetric sequence Eulerian) lengthens the pulse sequence and so increases the time T appearing in the Magnus expansion. Whether using this longer sequence actually improves the noise suppression depends on the values of the parameters $\epsilon\tau_0$, $\delta\tau_0$ and τ_0/t_0, but it could pay off if the pulse width is relatively large. Adding a gate pulse to the time-symmetric Eulerian memory sequence spoils the first-order decoupling and breaks the time symmetry, but the resulting contributions to Ω_1 and Ω_2 depend only on the width of the final pulse, not on the length of the pulse sequence.

Eulerian DD-protected gates that achieve exact first-order decoupling for nonzero-width pulses can be devised using the dynamically corrected gates recently introduced in [53, 54]. This scheme is based on the idea that distinct gates can have related errors, so that the errors cancel for a suitably constructed pulse sequence. The errors in distinct gates can be similar if the gates are constructed from control unitaries that traverse similar time-dependent baths, differing only by rescaling or reversing the time along the path. Arbitrary-order decoupling for nonzero-width pulses can be achieved by concatenating dynamically corrected gates [55].

VII. DERIVATIONS

In this section, we derive the coefficients for the bounds on the Magnus expansion listed in Table I. The Magnus expansion is computed for the Hamiltonian $H_M(t)$ given in Eq. (49); at any time t, $H_M(t) = \pm H_B + H'(t)$, where $H'(t)$ is either 0 or $\tilde{H}_{\rm err}(t)$. The two terms in $H_M(t)$ are bounded as $\|H_B\| \le \beta$ and $\|H'(t)\| \le J$; thus $\|H_M(t)\| \le \beta + J = \epsilon$. The Magnus terms can be computed from $H_M(t)$ using the following recursive formulas [56], derived in Appendix A:

$$A(t) = -iH_M(t); \tag{82a}$$

$$\Omega_1(T) = \int_0^T dt A(t); \tag{82b}$$

$$\Omega_n(T) = \sum_{j=1}^{n-1}\frac{B_j}{j!}\int_0^T dt S_n^{(j)}(t),\ \ n\ge 2; \tag{82c}$$

$$S_n^{(1)}(t) = \left[\Omega_{n-1}(t), A(t)\right]; \tag{82d}$$

$$S_n^{(j)}(t) = \sum_{m=1}^{n-j}\left[\Omega_m(t), S_{n-m}^{(j-1)}(t)\right], 2\le j\le n-1, \tag{82e}$$

where $\{B_j\}$ are the Bernoulli numbers. Explicit formulas for $\Omega_2(T)$ and $\Omega_3(T)$ were given in Eqs. (27) and (28).

A. General case

For the general (*i.e.*, not time-symmetric) case, Table I gives $C_1 = N\delta/T$ for regularly spaced pulses or $2N\delta/T$ in general, $C_2 = 1$, $C_3 = 4/9$, $C_4 = 11/9$ and $C_5 = 9.43$. Now we derive these coefficients.

1. *Bound for* Ω_1'

We assume that first-order decoupling is attained, so that in the limit of zero-width pulses $\tilde{H}(t)$ for the memory sequence satisfies Eq. (31): $\int_0^{t_{\rm DD}} dt \tilde{H}_{\rm err,0}(t) = 0$. Recall that the subscript "0" on $\tilde{H}_{\rm err}$ means we are to take δ to zero in $\tilde{H}_{\rm err}(t)$ while holding τ_0 fixed. If a zero-width gate pulse is appended to the memory sequence, then $\tilde{H}_{\rm err,0}(t)$ in the DD-protected gate differs from $\tilde{H}_{\rm err,0}(t)$ in the memory sequence only during the final instantaneous pulse, and therefore still integrates to zero. Hence the DD-protected gate as well as the memory sequence satisfies $\Omega_1(T) = -iTH_B$ and $\Omega_1'(T) = 0$.

When the pulses have finite width ($\delta > 0$), Ω_1' picks up corrections that depend on δ. Noting that $\tilde{H}(t)$ differs from $\tilde{H}_0(t)$ only during the pulses, we write

$$\begin{aligned}\Omega_1'(T) &= -i\int_0^{t_0} dt\tilde{H}(t) + it_0 H_B \\ &= -i\int_0^{t_0} dt\tilde{H}_0(t) + it_0 H_B \\ &\quad + i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_0(t) - i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}(t) \\ &= i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_0(t) - i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}(t). \end{aligned} \tag{83}$$

Here, dt_{PW} indicates integration only over times within the pulses. Now, $\tilde{H}_0(t) = H_B + \tilde{H}_{\mathrm{err},0}(t)$, so for a sequence with N pulses (including the gate pulse), we have $i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_0(t) = iN\delta H_B + i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_{\mathrm{err},0}(t)$ and $-i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}(t) = -iN\delta H_B - i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_{\mathrm{err}}(t)$. The two $iN\delta H_B$ terms cancel, and we are left with

$$\Omega_1'(T) = i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_{\mathrm{err},0}(t) - i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_{\mathrm{err}}(t). \tag{84}$$

The second term can be upper bounded by $N\delta J$. For the first term, Eq. (38) tells us that for $\delta = 0$, $\tilde{H}_0(t) = \tilde{H}^{(k)} = H_B + \tilde{H}_{\mathrm{err}}^{(k)}$ for $t \in [s_k, s_{k+1})$. Hence, we have that $i\int_0^{t_0} dt_{\mathrm{PW}}\tilde{H}_{\mathrm{err},0}(t) = i\delta\sum_k \tilde{H}_{\mathrm{err}}^{(k)}$. Now, the first-order decoupling condition can be written as

$$\int_0^{t_0} dt\tilde{H}_{\mathrm{err},0} = \sum_k (s_{k+1} - s_k)\tilde{H}_{\mathrm{err}}^{(k)} = 0. \tag{85}$$

If all the pulses are regularly spaced in time, so that $s_{k+1} - s_k$ are all equal for all k, this condition implies that $\sum_k \tilde{H}_{\mathrm{err}}^{(k)} = 0$. In this case, the first term of the right-hand side of Eq. (84) vanishes and $\Omega_1'(T)$ is bounded by the norm of the second term only:

$$\|\Omega_1'(T)\| \le N\delta J = \frac{N\delta}{T}(JT). \tag{86}$$

Hence, $C_1 = N\delta/T$ if pulses are regularly spaced in time. Even if the pulses are not regularly spaced in time, this value of C_1 works whenever $\sum_k \tilde{H}_{\mathrm{err}}^{(k)} = 0$. Otherwise, we can still upper bound the first term in Eq. (84) by $N\delta J$, so that $\|\Omega_1'(T)\| \le 2N\delta J = (2N\delta/T)(JT)$. This gives $C_1 = 2N\delta/T$ in general.

2. *Bound for* Ω_2

To bound $\Omega_2(T)$, we use the formula Eq. (27) for $\Omega_2(T)$:

$$\Omega_2(T) = -\frac{1}{2}\int_0^T ds_1\int_0^{s_1} ds_2[H_M(s_1), H_M(s_2)]. \tag{87}$$

Since $H_M(t) = \pm H_B + H'(t)$, the integrand $[H_M(s_1), H_M(s_2)]$ splits into a sum of three nonzero terms: $\pm[H_B, H'(s_2)]$, $\pm[H'(s_1), H_B]$ and $[H'(s_1), H'(s_2)]$. These are respectively bounded by $2J\beta$, $2J\beta$ and $2J^2$, which gives for all $s_1, s_2 \in [0, T]$,

$$\|[H_M(s_1), H_M(s_2)]\| \le 2J(2\beta + J) \le 4J\epsilon. \tag{88}$$

Putting this into the norm of Eq. (87) and doing the integrals, we get

$$\|\Omega_2(T)\| \le 2J\epsilon\int_0^T ds_1\int_0^{s_1} ds_2(1) = (JT)(\epsilon T), \tag{89}$$

which gives $C_2 = 1$.

3. *Bound for* Ω_3

For $\Omega_3(T)$, we start from the explicit formula Eq. (28). It is convenient to have a more compact notation that will also be useful in the time-symmetric case later. With $A(t) \equiv -iH_M(t)$, let us denote $A(s_i)$ (occurring in the integrand of Eq. (28)) by its index i, and let $[ijk] \equiv [A(s_i), [A(s_j), A(s_k)]]$. Furthermore, let $\Theta(x)$ be the step function: $\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$, and define

$$\Theta(pq) \equiv \Theta(t_p - t_q), \quad \Theta(pqr) \equiv \Theta(pq)\Theta(qr). \tag{90}$$

Then $\Omega_3(T)$ can be written compactly as

$$\begin{aligned}\Omega_3(T) = \frac{1}{6}\int_0^T ds_1\int_0^T ds_2\int_0^T ds_3\Theta(123) \\ \times\left([123] + [321]\right). \end{aligned} \tag{91}$$

We need a bound on the triple commutator $[ijk]$:

$$\|[ijk]\| \le 2\epsilon\|[jk]\| \le 2\epsilon(4J\epsilon) = 8J\epsilon^2 \tag{92}$$

for any i, j, k. Therefore we find

$$\begin{aligned}\|\Omega_3(T)\| &\le \frac{1}{6}\left(16J\epsilon^2\right)\int_0^T ds_1 ds_2 ds_3\Theta(123) \\ &= \frac{8}{3}J\epsilon^2\left(\frac{T^3}{6}\right) \\ &\le \frac{4}{9}(JT)(\epsilon T)^2. \end{aligned} \tag{93}$$

Hence, $C_3 = 4/9$.

4. *Bounds for* $\Omega_{n\ge 4}$

To bound the Magnus terms for $n \ge 4$, we use the recursive formulas Eq. (82) and ideas from [57, 58]. In Appendix E, we show that the $S_n^{(j)}$ operators satisfy:

$$\|S_n^{(j)}(t)\| \le f_n^{(j)} J\left(2\epsilon t\right)^{n-1}, \tag{94}$$

for all $n \geq 2$, $1 \leq j \leq n-1$, where the coefficients $f_n^{(j)}$ are given in Eq. (E2). Using this, we can write down bounds for $\Omega_{n\geq 4}$ as follows:

$$\begin{aligned}\|\Omega_n(T)\| &\leq \sum_{j=1}^{n-1} \frac{|B_j|}{j!} \int_0^T ds \|S_n^{(j)}(s)\| \\ &\leq \frac{1}{n} \sum_{j=1}^{n-1} \frac{|B_j|}{j!} f_n^{(j)} (JT)(2\epsilon T)^{n-1} \\ &= f_n(JT)(4\epsilon T)^{n-1}, \end{aligned} \tag{95}$$

where the coefficients f_n are defined as

$$f_n = \frac{1}{n 2^{n-1}} \sum_{j=1}^{n-1} \frac{|B_j|}{j!} f_n^{(j)}. \tag{96}$$

Using Eq. (96) and the recursive formula for $f_n^{(j)}$ from Eq. (E2), one can show that $f_4 = 11/576$. Then, $\Omega_4(T)$ can be bounded as

$$\|\Omega_4(T)\| \leq \frac{11}{576}(JT)(4\epsilon T)^3, \tag{97}$$

so $C_4 = 4^3(11/576) = 11/9$.

The bounds for Ω_n for $n \geq 5$ can be gathered together into a single bound by writing

$$\sum_{n=5}^{\infty} \|\Omega_n(T)\| \leq (JT)(4\epsilon T)^4 \left[\sum_{n=5}^{\infty} f_n (4\epsilon T)^{n-5} \right]. \tag{98}$$

In [58], the $\{f_n\}$ were shown to be coefficients in the power series expansion of $G^{-1}(y) = \sum_{n=1}^{\infty} f_n y^n$; $G^{-1}(y)$ is the inverse function of

$$y = G(s) = \int_0^s dx \left[2 + \frac{x}{2}\left(1 - \cot\frac{x}{2}\right)\right]^{-1}, \tag{99}$$

defined for domain $-2\pi \leq s \leq 2\pi$, the interval over which $G(s)$ is monotonically increasing. A self-contained proof of this fact is provided in Appendix F. We want to relate the expression in the brackets in Eq. (98) to G^{-1}. Define ζ as

$$\zeta = G(2\pi) = 2.17374\ldots, \quad G^{-1}(\zeta) = 2\pi, \tag{100}$$

and assume that $\epsilon T \leq 0.54$ so that $4\epsilon T \leq \zeta$. Then, $G^{-1}(4\epsilon T) \leq 2\pi$ since $G(s)$ is monotonically increasing over its domain, and therefore,

$$\begin{aligned}\left[\sum_{n=5}^{\infty} f_n (4\epsilon T)^{n-5}\right] &\leq \sum_{n=5}^{\infty} f_n \zeta^{n-5} \\ &= \frac{1}{\zeta^5}\left[G^{-1}(\zeta) - \sum_{n=1}^{4} f_n \zeta^n \right]. \end{aligned} \tag{101}$$

Using $f_1 = 1$, $f_2 = \frac{1}{4}$, $f_3 = \frac{5}{72}$ and $f_4 = \frac{11}{576}$, which can be derived from Eq. (96) and Eq. (E2), Eq. (101) implies

$$\left[\sum_{n=5}^{\infty} f_n (4\epsilon T)^{n-5}\right] \leq 0.03685\ldots \equiv C'. \tag{102}$$

Then,

$$\sum_{n=5}^{\infty} \|\Omega_n(T)\| \leq C'(JT)(4\epsilon T)^4. \tag{103}$$

Therefore, $C_5 \equiv 4^4 \times C' \simeq 9.43$.

Note that the condition $\epsilon T \leq 0.54$ is more stringent than the sufficient condition for convergence of the Magnus expansion given in Eq. (29), which requires $\epsilon T < \pi$. If $0.54 < \epsilon T < \pi$, we need to use a different method to find an upper bound on the sum of the high-order Magnus terms.

B. Time-symmetric case

Now we derive bounds on the Magnus terms that apply when the pulse sequence is time-symmetric except inside a small region Δ; as before, we use Δ to denote both this region and its size. As in Eq. (49), we are interested in the Hamiltonian $H_M(t)$ describing evolution for time Γ governed by the Hamiltonian $-H_B$, followed by evolution for time $T-\Gamma$ governed by the toggling-frame Hamiltonian of a DD-protected gate. But our analysis in this Section applies to any Hamiltonian $H_M(t)$ that is time-symmetric outside region Δ.

We bound the first-order term $\Omega_1'(T)$ just as in the general case, since time symmetry makes no difference for this term. But even Magnus terms vanish when $\Delta = 0$, and we will derive explicit Δ-dependent bounds on $\Omega_2(T)$ and $\Omega_4(T)$. We bound $\Omega_3(T)$ starting from Eq. (91), but exploit the time symmetry to improve on the bound that applies in the general case. We could also exploit the time symmetry to derive improved bounds on the higher-order Magnus terms ($\Omega_{n\geq 5}(T)$); however we will not bother to do so. Instead we use the same upper bounds on these terms that apply in the general case, with the expectation that these bounds are already quite small in typical cases of interest.

1. *Bounds for Ω_2 and Ω_4*

To obtain a bound on $\Omega_2(T)$ for a nearly time-symmetric sequence, we note that the double time integral in Eq. (87) can be split into four cases: (i) $t_1, t_2 \notin \Delta$, (ii) $t_1 \in \Delta, t_2 \notin \Delta$, (iii) $t_1 \notin \Delta, t_2 \in \Delta$ and (iv) $t_1, t_2 \in \Delta$. The contribution from Case (i) vanishes, because $H_M(t)$ is time-symmetric in this region. The remaining cases can be bounded by first bounding the integrand (the commutator) and then doing the integration. The commutator can be bounded as in Eq. (88): $\|[H_M(t_1), H_M(t_2)]\| \leq 4J\epsilon$. Then doing the integration gives

$$\begin{aligned}\|\Omega_2(T)\| &\leq \frac{1}{2}(4J\epsilon)\left[2\Delta(T-\Delta) + \Delta^2\right] \\ &= 4\frac{\Delta}{T}\left(1 - \frac{\Delta}{2T}\right)(JT)(\epsilon T), \end{aligned} \tag{104}$$

which vanishes as expected in the limit $\Delta \to 0$. Hence, $C_2 = 4(\Delta/T)\,(1-\Delta/2T)$.

Similarly, we can derive upper bounds on all even Magnus terms that depend linearly on Δ/T to lowest order, as discussed in Appendix D. Eq. (D7) gives the same bound on $\Omega_2(T)$ as Eq. (104), and for $\Omega_4(T)$ yields

$$\begin{aligned}\|\Omega_4(T)\| &\le 14(JT)(\epsilon T)^3\left[1-(1-\frac{\Delta}{T})^4\right]\\ &= 14(JT)(\epsilon T)^3\left[4\left(\frac{\Delta}{T}\right)-6\left(\frac{\Delta}{T}\right)^2\right.\\ &\qquad\left.+4\left(\frac{\Delta}{T}\right)^3-\left(\frac{\Delta}{T}\right)^4\right]. \qquad (105)\end{aligned}$$

Since $4(\Delta/T)^3 \le 4(\Delta/T)^2$ and $(\Delta/T)^4 \ge 0$, we can rewrite this as

$$\begin{aligned}\|\Omega_4(T)\| &\le 14(JT)(\epsilon T)^3\left[4\frac{\Delta}{T}-2\left(\frac{\Delta}{T}\right)^2\right]\\ &= 56\frac{\Delta}{T}\left(1-\frac{\Delta}{2T}\right)(JT)(\epsilon T)^3. \qquad (106)\end{aligned}$$

Hence, $C_4 = 56(\Delta/T)\,(1-\Delta/2T)$.

2. Bound for Ω_3

To bound $\Omega_3(T)$ in the time-symmetric case, we begin with Eq. (91). Let us first assume that $H_M(t)$ is perfectly time-symmetric, so $\Delta = 0$. In this case, by dividing the interval $[0,T]$ into two half-intervals $[0,T/2]$ and $[T/2,T]$ and using the time-symmetry, we find that we can reduce the range of integration for t_1, t_2 and t_3 in Eq. (91) to $[0,T/2]$ while making the replacement

$$\Theta(123) \to \Theta(123) + \Theta(321) + \Theta(23) + \Theta(21) \qquad (107)$$

The four terms in the replacement arise from the four cases: (i) $t_1,t_2,t_3 \in [0,T/2]$, (ii) $t_1,t_2,t_3 \in [T/2,T]$, (iii) $t_1 \in [T/2,T]$ and $t_2,t_3 \in [0,T/2]$, and (iv) $t_1,t_2 \in [T/2,T]$ and $t_3 \in [0,T/2]$. These are the only cases for which $\Theta(123)$ is nonvanishing. Hence, we have

$$\begin{aligned}\Omega_3(T) &= \frac{1}{6}\int_0^{T/2} dt_1dt_2dt_3\,([123]+[321])\\ &\qquad\times(\Theta(123)+\Theta(321)+\Theta(23)+\Theta(21))\\ &= \frac{1}{3}\int_0^{T/2} dt_1dt_2dt_3[123] \qquad (108)\\ &\qquad\times(\Theta(123)+\Theta(321)+\Theta(23)+\Theta(21))\end{aligned}$$

where in the second equality, we have changed variables $1 \leftrightarrow 3$ in the term containing $[321]$.

Next, observe that

$$\begin{aligned}\Theta(23) &= \Theta(123)+\Theta(213)+\Theta(231),\\ \Theta(21) &= \Theta(321)+\Theta(231)+\Theta(213), \qquad (109)\end{aligned}$$

and note that

$$0 = \int_0^{T/2} dt_1dt_2dt_3\,(\Theta(321)+\Theta(231))\,[123], \qquad (110)$$

since $[123]$ is antisymmetric under the replacement $2 \leftrightarrow 3$ while $\Theta(321)+\Theta(231)$ is symmetric. Thus, we obtain

$$\Omega_3(T) = \frac{1}{3}\int_0^{T/2} dt_1dt_2dt_3\,(2\Theta(123)+2\Theta(213))\,[123]. \qquad (111)$$

Bounding the norm of $[123]$ using Eq. (92), and then changing variables $1 \leftrightarrow 2$ in the term with $\Theta(213)$, we finally have that

$$\begin{aligned}\|\Omega_3(T)\| &\le \frac{4}{3}\left(8J\epsilon^2\right)\int_0^{T/2} dt_1dt_2dt_3\Theta(123)\\ &= \frac{4}{3}\left(8J\epsilon^2\right)\left(\frac{T^3}{48}\right)\\ &= \frac{2}{9}(JT)(\epsilon T)^2, \qquad (112)\end{aligned}$$

for a perfectly time-symmetric $H_M(t)$. Thus $C_3 = 2/9$ for $\Delta = 0$. Observe that Eq. (112) has an additional factor of $1/2$ compared to the corresponding bound (Eq. (93)) in the general non-time-symmetric case.

If $H_M(t)$ is time-symmetric except for $t \in \Delta$, we can do an analysis similar to what we did for the even Magnus terms. Let $H'_M(t) = H_M(t)$ for $t \notin \Delta$ and $H'_M(t) = 0$ for $t \in \Delta$. Then, $H'_M(t)$ is perfectly time-symmetric. Actually, we can discard the Δ time period for which $H'_M(t) = 0$ and glue the pieces together (shifting the time t accordingly) so that we end up with a perfectly time-symmetric $H''_M(t)$ for $t \in [0, T-\Delta]$. Then, we can rewrite the expression for $\Omega_3(T)$ given in Eq. (91) as

$$\begin{aligned}\Omega_3(T) &= \frac{1}{6}\int_0^{T} ds_1ds_2ds_3\Theta(123)\,([123]+[321])\\ &= \frac{1}{6}\int (ds_1ds_2ds_3)_{T\setminus\Delta}\Theta(123)\,([123]+[321])\\ &\quad+\frac{1}{6}\int\,(ds_1ds_2ds_3)_{\Delta}\,\Theta(123)\,([123]+[321])\,, \qquad (113)\end{aligned}$$

where $(\cdot)_{T\setminus\Delta}$ denotes the condition that none of the integration variables in the parentheses is in Δ, while $(\cdot)_\Delta$ denotes at least one of the integration variables in the parentheses is in Δ. The first integral in Eq. (113) is just $\Omega_3(T-\Delta)$ computed for the perfectly time-symmetric $H''_M(t)$, and hence can be bounded using Eq. (112). The second integral can be bounded by first bounding the commutators using Eq. (92), and then doing the time-integral $\int(ds_1ds_2ds_3)_\Delta\Theta(123)$. This time-integral splits into three cases: (i) exactly one of s_1,s_2,s_3 is in Δ, (ii) exactly two of s_1,s_2,s_3 are in Δ, and (iii) all three of s_1,s_2,s_3 are in Δ.

In case (i), for $s_1 \in \Delta$ and $s_2, s_3 \notin \Delta$, the time-integral is given by

$$\begin{aligned}
&\int (ds_1)_\Delta \int (ds_2 ds_3)_{T\backslash\Delta} \Theta(123) \\
&\leq \int (ds_1)_\Delta \int (ds_2 ds_3)_{T\backslash\Delta} \Theta(23) \\
&= \Delta \int_0^{T-\Delta} ds_2 ds_3 \Theta(23) \\
&= \frac{1}{2}\Delta(T-\Delta)^2,
\end{aligned} \tag{114}$$

where in going from the second to the third line, we have evaluated the s_1 integral, and glued the different pieces in $T\backslash\Delta$ into a single continuous time interval $[0, T-\Delta]$. The time-integrals for the other possibilities in case (i), i.e., ($s_2 \in \Delta, s_1, s_3 \notin \Delta$) and ($s_3 \in \Delta, s_1, s_2 \notin \Delta$), can be bounded in the same way.

In case (ii), for $s_1, s_2 \in \Delta$ and $s_3 \notin \Delta$, a similar argument as used in Eq. (114) can be used to bound the time-integral:

$$\begin{aligned}
&\int (ds_1)_\Delta \int (ds_2)_\Delta \int (ds_3)_{T\backslash\Delta} \Theta(123) \\
&\leq (T-\Delta) \int_0^{\Delta} ds_1 ds_2 \Theta(12) \\
&= \frac{1}{2}\Delta^2(T-\Delta).
\end{aligned} \tag{115}$$

The remaining two possibilities in case (ii), i.e., ($s_1, s_3 \in \Delta, s_2 \notin \Delta$) and ($s_2, s_3 \in \Delta, s_1 \notin \Delta$), can be bounded in the same way.

In case (iii), we only have one possibility: $s_1, s_2, s_3 \in \Delta$. The corresponding time-integral can be computed as

$$\begin{aligned}
&\int (ds_1)_\Delta \int (ds_2)_\Delta \int (ds_3)_\Delta \Theta(123) \\
&= \int_0^{\Delta} ds_1 ds_2 ds_3 \Theta(123) \\
&= \frac{1}{6}\Delta^3.
\end{aligned} \tag{116}$$

Finally, we have that

$$\begin{aligned}
&\|\Omega_3(T)\| \\
&\leq \frac{2}{9}\left[J(T-\Delta)\right]\left[\epsilon(T-\Delta)\right]^2 + \frac{1}{3}(8J\epsilon^2) \\
&\quad \times \left[\frac{3}{2}\Delta(T-\Delta)^2 + \frac{3}{2}\Delta^2(T-\Delta) + \frac{1}{6}\Delta^3\right] \\
&= \frac{2}{9}(JT)(\epsilon T)^2\left(1 + 15\frac{\Delta}{T} - 15\frac{\Delta^2}{T^2} + \frac{\Delta^3}{T^3}\right) \\
&\leq \frac{2}{9}(JT)(\epsilon T)^2\left\{1 + 15\left(\frac{\Delta}{T}\right)\left[1 - \frac{14}{15}\left(\frac{\Delta}{T}\right)\right]\right\}
\end{aligned} \tag{117}$$

where in the last line we have used $\Delta^3/T^3 \leq \Delta^2/T^2$. This yields C_3 for the time-symmetric case as stated in Table I.

VIII. CONCATENATED DYNAMICAL DECOUPLING

A concatenated DD pulse sequence is a recursively generated sequence with a self-similar structure [9, 10]. For example, from the "level-1" universal pulse sequence

$$p_1 = Z\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I} \tag{118}$$

we obtain the corresponding "level-2" sequence by replacing each pulse interval $\mathbb{I}$ in the level-1 sequence by the complete level-1 sequence p_1, obtaining

$$p_2 = Zp_1Xp_1Zp_1Xp_1; \tag{119}$$

similarly, the level-k sequence is

$$p_k = Zp_{k-1}Xp_{k-1}Zp_{k-1}Xp_{k-1}. \tag{120}$$

If the duration of a single pulse is τ_0 and p_1 is an R-pulse sequence that achieves first-order decoupling, then the corresponding level-k sequence p_k has duration $T^{(k)} = R^k\tau_0$ and achieves kth-order decoupling; *i.e.*, has effective noise strength $O(J\epsilon^k)$. Concatenated pulse sequences are less efficient than sequences with nonuniform pulse intervals that achieve kth-order decoupling with fewer pulses, but nevertheless have some nice properties. For one thing, the concatenation method can be applied to enhance any pulse sequence that achieves first-order (or higher-order) decoupling, while "optimal" pulse sequences [13, 39, 41, 59] are known only for qubits. Furthermore, concatenated pulse sequences are relatively robust against pulse imperfections, because pulse errors arising at each level get suppressed at higher levels. We will analyze the performance of concatenated pulse sequences in two ways, first using the Magnus expansion, and then in Sec. X E using the Dyson expansion and bath correlation functions.

A. Magnus expansion analysis

The noise Hamiltonian has an unambiguous decomposition into two parts: $H = H_B + H_{\rm err}$, where $H_B = I \otimes B_0$, $H_{\rm err} = \sum_\alpha S_\alpha \otimes B_\alpha$, and $\{S_\alpha\}$ is a basis for the *traceless* operators acting on the system. For a level-1 pulse sequence with duration T, the toggling-frame time evolution operator is $\tilde{U}(T) = \exp(\Omega(T))$; writing $\Omega(T) = -iH^{(1)}T$, we may regard $H^{(1)}$ as the level-1 "effective Hamiltonian." Like H, $H^{(1)}$ has an unambiguous decomposition into two parts,

$$H^{(1)} = H_B^{(1)} + H_{\rm err}^{(1)} \equiv I \otimes B_0^{(1)} + \sum_\alpha S_\alpha \otimes B_\alpha^{(1)}, \tag{121}$$

and we may define parameters that characterize the effective noise at level 1:

$$\|H_B^{(1)}\| \leq \beta^{(1)}, \quad \|H_{\rm err}^{(1)}\| \leq J^{(1)}, \quad \epsilon^{(1)} = \beta^{(1)} + J^{(1)}. \tag{122}$$

Now, we can analyze the level-2 pulse sequence just as we did the level-1 sequence, but with the level-0 noise Hamiltonian $H = H^{(0)}$ replaced by the level-1 effective Hamiltonian $H^{(1)}$. Proceeding in this way, we can estimate properties of the toggling-frame time evolution operator $\widetilde{U}^{(k)}$ for the level-k pulse sequence using the level-$(k-1)$ Hamiltonian $H^{(k-1)}$. At each level, we can define noise parameters $\beta^{(k)}$, $J^{(k)}$, and $\epsilon^{(k)}$ as in Eq. (122), and derive recursion relations that relate the level-k noise parameters to level-$(k-1)$ noise parameters.

To understand how first-order decoupling is achieved by the level-1 sequence, we assumed that the toggling-frame Hamiltonian is constant in the interval between pulses. For the concatenated sequence at level 2 and above, this assumption is not true, since the interval in between the level-k pulses contains a complex level-$(k-1)$ pulse sequence. However the unitary operator describing the evolution from the end of one level-k pulse to the beginning of the next level-k pulse is equivalent to the evolution operator that would have been derived from the constant Hamiltonian $H^{(k-1)}$ during the pulse interval. Thus for the purpose of understanding the time evolution in the toggling frame resulting from the level-k sequence, it does no harm to imagine that the Hamiltonian is constant between pulses and do the analysis just as for the level-1 sequence.

For a sequence that achieves first-order decoupling, Ω_1 at each level is a pure bath term

$$\Omega_1^{(k)}(T^{(k)}) = -iT^{(k)} H_B^{(k-1)}, \tag{123}$$

where $T^{(k)} = R^k \tau_0$ is the duration of the level-k sequence, constructed by concatenating k times a sequence with R pulses. Suppose we consider a pulse sequence such that each pulse either commutes or anticommutes with each of the traceless operators in the set $\{S_\alpha\}$ (the argument below can be easily adapted to more general pulse sequences). Under this assumption, the second-order term Ω_2 in the Magnus expansion has no pure-bath component (see Appendix G) and thus contributes only to $H_{\rm err}^{(1)}$. Therefore $H_B^{(k)}$ arises from $\Omega_1^{(k)}$ and the pure bath component of $\Omega_{\geq 3}^{(k)} = \sum_{n=3}^{\infty} \Omega_n^{(k)}$. As shown in Appendix H, the norm of the pure-bath component of $\Omega_j^{(k)}$ is no larger than $\|\Omega_j^{(k)}\|$; it follows that we may choose $\beta^{(k)}$ such that $\beta^{(k)} T^{(k)}$ is an upper bound on

$$\|\Omega_1^{(k)}(T)\| + \|\Omega_{\geq 3}^{(k)}(T)\|. \tag{124}$$

From Eq. (50) and Table I we see that

$$\|\Omega_{\geq 3}^{(k)}\| \leq c_3^{(k)} \left(J^{(k-1)} T^{(k)}\right) \left(\epsilon^{(k-1)} T^{(k)}\right)^2, \tag{125}$$

where the "constant" $c_3^{(k)}$ actually depends on the value of $\epsilon^{(k-1)} T^{(k)}$:

$$c_3^{(k)} = \frac{4}{9} + \frac{11}{9}\left(\epsilon^{(k-1)} T^{(k)}\right) + 9.43\left(\epsilon^{(k-1)} T^{(k)}\right)^2, \tag{126}$$

assuming $\epsilon^{(k-1)} T^{(k)} \leq .54$ (*e.g.*, $c_3^{(k)} = 0.67$ for $\epsilon^{(k-1)} T^{(k)} = 0.1$ and $c_3^{(k)} = 0.46$ for $\epsilon^{(k-1)} T^{(k)} = 0.01$). Recalling Eq. (123), we conclude that

$$\beta^{(k)} = \beta^{(k-1)} + c_3^{(k)} J^{(k-1)} \left(\epsilon^{(k-1)} T^{(k)}\right)^2. \tag{127}$$

Though Eq. (127) has been expressed as an equality, the right-hand side is actually an upper bound on $\|H_B^{(k)}\|$.

The level-k error Hamiltonian $H_{\rm err}^{(k)}$ arises from $\Omega_2^{(k)}$ and the traceless component of $\Omega_{\geq 3}^{(k)}$. It is shown in Appendix H that the norm of the traceless component of $\Omega_j^{(k)}$ is no larger than $2\|\Omega_j^{(k)}\|$; therefore we may choose $J^{(k)}$ such that $J^{(k)} T^{(k)}$ is an upper bound on

$$\|\Omega_2^{(k)}(T)\| + 2\|\Omega_{\geq 3}^{(k)}(T)\|. \tag{128}$$

(If the system is a single qubit, then the norm of the traceless component of $\Omega_j^{(k)}$ is no larger than $\|\Omega_j^{(k)}\|$, and thus the factor of 2 in the second term can be omitted.) Therefore, again using Eq. (50) and Table I we find

$$J^{(k)} = c_2^{(k)} J^{(k-1)} \left(\epsilon^{(k-1)} T^{(k)}\right), \tag{129}$$

where

$$\begin{aligned} c_2^{(k)} &= 1 + 2\left[\frac{4}{9}\left(\epsilon^{(k-1)} T^{(k)}\right) + \frac{11}{9}\left(\epsilon^{(k-1)} T^{(k)}\right)^2 \right. \\ &\qquad \left. + 9.43\left(\epsilon^{(k-1)} T^{(k)}\right)^3\right], \end{aligned} \tag{130}$$

assuming $\epsilon^{(k-1)} T^{(k)} \leq .54$ (*e.g.*, $c_2^{(k)} = 1.133$ for $\epsilon^{(k-1)} T^{(k)} = 0.1$ and $c_2^{(k)} = 1.01$ for $\epsilon^{(k-1)} T^{(k)} = 0.01$).

Eq. (127) can be rewritten as

$$\beta^{(k)} = \beta^{(k-1)} + K^{(k-1)}. \tag{131}$$

where

$$K^{(k-1)} = c_3^{(k)} \left(J^{(k-1)} T^{(k)}\right) \left(\epsilon^{(k-1)} T^{(k)}\right)^2, \tag{132}$$

and iterating this equation yields

$$\beta^{(k)} = \beta + K^{(1)} + K^{(2)} + \cdots + K^{(k-1)} \tag{133}$$

and

$$\begin{aligned} \epsilon^{(k)} &= \beta^{(k)} + J^{(k)} \\ &= \beta + K^{(1)} + K^{(2)} + \cdots + K^{(k-1)} + J^{(k)}. \end{aligned} \tag{134}$$

The solution to the recursion relations Eq. (129), (132), (134) cannot be expressed easily in closed form, but the properties of the solution can be grasped if we assume that

$$c_2^{(k)} \epsilon^{(k-1)} \leq \bar{c}\epsilon \tag{135}$$

for each k, where $\bar{c}$ is a constant. That is, if we iterate the recursion relations to estimate $J^{(\ell)}$, our assumption is that Eq. (135) is satisfied for all $k \le \ell$. Then using $T^{(k)} = R^k \tau_0$, we can replace Eq. (129) by

$$J^{(k)} = (\bar{c}\epsilon\tau_0)\, R^k J^{(k-1)}, \tag{136}$$

which has the solution

$$J^{(k)} = (\bar{c}\epsilon\tau_0)^k\, R^{k(k+1)/2} J \tag{137}$$

where $J^{(0)} = J$ [60].

The effective noise strength for the level-k sequence is

$$\begin{aligned} \eta_{\rm DD}^{(k)} &= \|H_{\rm err}^{(k)}\| T^{(k)} = J^{(k)} R^k \tau_0 \\ &= R^{k(k+3)/2}\, (\bar{c}\epsilon\tau_0)^k\, (J\tau_0), \end{aligned} \tag{138}$$

so that

$$\eta_{\rm DD}^{(k)} = R^{k+1} (\bar{c}\epsilon\tau_0) \eta_{\rm DD}^{(k-1)}; \tag{139}$$

therefore the optimal suppression of the noise strength is achieved by choosing the level k to be the largest integer such that $R^{k+1}(\bar{c}\epsilon\tau_0) < 1$, or equivalently,

$$k_{\rm max} = \lfloor \log_R(1/\bar{c}\epsilon\tau_0) - 1 \rfloor. \tag{140}$$

where $\lfloor\cdot\rfloor$ denotes the "floor" function. For example, if $\bar{c}\epsilon\tau_0 = 10^{-3}$ and $R = 4$, we choose $k_{\rm max} = 3$ (*i.e.*, a sequence with duration $T^{(3)} = 64\tau_0$) and obtain $\eta_{\rm DD}/(J\tau_0) = 2.6\times 10^{-4}$, an improvement by a factor of 60 over the noise strength achieved by the level-1 sequence.

The expression for $\eta_{\rm DD}^{(k)}$ in Eq. (138) is the exponential of a quadratic function of k, minimized at $k = \log(1/\bar{c}\epsilon\tau_0) - 3/2$. The nearest integer differs from this optimal value by at most 1/2; substituting $k + 1 = \log(1/\bar{c}\epsilon\tau_0)$ into Eq. (138), we conclude that the optimal effective noise strength satisfies

$$\eta_{\rm DD}^{(\rm opt)}/(J\tau_0) \le R^{-1}\, (\bar{c}\epsilon\tau_0)^{\frac{1}{2}\log_R(1/\bar{c}\epsilon\tau_0) - \frac{3}{2}}. \tag{141}$$

If we concatenate a time-symmetric pulse sequence, which achieves second-order decoupling, then we may replace Eq. (129) by

$$J^{(k)} = 2c_3^{(k)} J^{(k-1)} \left(\epsilon^{(k-1)} T^{(k)}\right)^2 \tag{142}$$

(the factor of 2 can be omitted if the system is a qubit), and we can also improve the estimate of c_3 to

$$c_3^{(k)} = \frac{2}{9} + 9.43 (\epsilon^{(k-1)} T^{(k)})^2, \tag{143}$$

where $\epsilon^{(k-1)} T^{(k)} \le .54$. Defining $\bar{c}$ for a time-symmetric sequence by $2c_3^{(k)} \left(\epsilon^{(k-1)}\right)^2 \le (\bar{c}\epsilon)^2$, Eq. (142) becomes

$$J^{(k)} = (\bar{c}\epsilon\tau_0)^2\, R^{2k} J^{(k-1)}, \tag{144}$$

which has the solution

$$J^{(k)} = (\bar{c}\epsilon\tau_0)^{2k}\, R^{k(k+1)} J, \tag{145}$$

and thus

$$\eta_{\rm DD}^{(k)} = R^{k(k+2)}\, (\bar{c}\epsilon\tau_0)^{2k}\, (J\tau_0). \tag{146}$$

The noise strength is optimized by choosing the largest integer k such that $k + \frac{1}{2} < \log_R(1/\bar{c}\epsilon\tau_0)$. For example, if $R = 8$ and $\bar{c}\epsilon\tau_0 = 10^{-3}$, we choose $k = 2$ (*i.e.*, a sequence with duration $T^{(2)} = 64\tau_0$) and obtain $\eta_{\rm DD}/(J\tau_0) = 1.7\times 10^{-5}$, an improvement by a factor of 30 over the noise strength achieved by the level-1 sequence. The optimal noise strength satisfies

$$\eta_{\rm DD}^{(\rm opt)}/(J\tau_0) \le R^{-3/4}\, (\bar{c}\epsilon\tau_0)^{\log_R(1/\bar{c}\epsilon\tau_0) - 2}. \tag{147}$$

Optimal noise strengths for the universal and time-symmetric sequences, plotted in Fig. 8, are orders of magnitude lower than the noise strengths achievable without concatenation, shown in Fig. 5. Though longer, the time-symmetric sequence performs much better when $\bar{c}\epsilon\tau_0$ is sufficiently small.

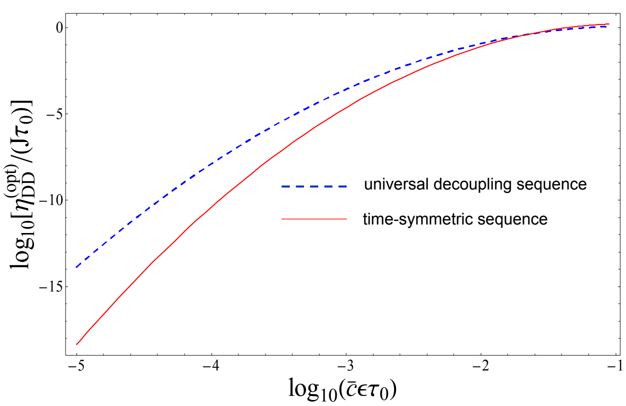

FIG. 8: Upper bounds on the effective noise strengths achieved by the concatenated universal DD pulse sequence (blue dashed line, based on Eq. (141) with $R = 4$) and by the concatenated time-symmetric DD pulse sequence (red solid line, based on Eq. (147) with $R = 8$), as a function of $\bar{c}\epsilon\tau_0$, where $\bar{c}$ is defined in the text.

B. Including pulse errors

How is this analysis affected if the pulses are imperfect? The answer depends on the degree to which the pulse errors are systematic and reproducible, rather than random. As in our discussion of Eulerian decoupling, let us assume that the errors are systematic. This assumption is reasonable if the pulse errors arise from the time-independent noise Hamiltonian that is "on" during the pulses, rather than from variations in the pulse shape.

In the recursive analysis of the concatenated pulse sequence, the effective Hamiltonian $H^{(k-1)}$ incorporates all the damage caused by the pulses errors at level $k-1$ and below. Because the pulse errors are systematic, we may use the same $H^{(k-1)}$ to describe the noise in each interval between level-k pulses. Suppose we imagine, at first, that while the pulses at level $k-1$ and below are noisy, the pulses at level k are ideal, and denote by $\hat{J}^{(k)}$ the upper bound on $\|H_{\rm err}^{(k)}\|$ under this fictitious assumption. Repeating the derivation of Eq. (129) yields

$$\hat{J}^{(k)} = c_2^{(k)} J^{(k-1)} \left(\epsilon^{(k-1)} T^{(k)}\right). \tag{148}$$

But now we must relate $\hat{J}^{(k)}$ to $J^{(k)}$ by estimating the effects of the pulse errors at the top level.

The noise in these level-k pulses is governed by the level-0 error Hamiltonian $H_{\rm err}^{(0)}$ rather than the effective level-$(k-1)$ error Hamiltonian $H_{\rm err}^{(k-1)}$. We could adapt our analysis of the Magnus expansion to this new situation, using a different upper bound on $H_{\rm err}$ during the pulses than in the interval between pulses, but then we would face the complication of revising our estimate of all the higher-order terms in the expansion. To avoid that complication, we use a different approach. As in Sec. VII A 1, we assume that the Hamiltonian describing the sequence of noisy pulses at level k deviates in operator norm from the Hamiltonian describing the sequence of ideal pulses at level k by at most $2J$ during a total time interval $R\delta$, if there are R pulses each with width δ. It then follows from Lemma 3 in Appendix C that

$$\left\| e^{\Omega^{(k)}} - e^{\hat{\Omega}^{(k)}} \right\| \le 2R\delta J, \tag{149}$$

where $\Omega^{(k)}$ includes pulse-error corrections at all levels while $\hat{\Omega}^{(k)}$ includes pulse-error corrections at level $k-1$ and below but not at level k. From Eq. (I9) in Appendix I, we find that

$$\left\| \Omega^{(k)} - \hat{\Omega}^{(k)} \right\| \le d^{(k)} \left\| e^{\Omega^{(k)}} - e^{\hat{\Omega}^{(k)}} \right\| \le 2d^{(k)} R\delta J, \tag{150}$$

where the "constant" $d^{(k)}$ is close to one if $\|\Omega^{(k)}\|$ and $\|\hat{\Omega}^{(k)}\|$ are both small; therefore we obtain an upper bound on $J^{(k)}$:

$$\begin{aligned} J^{(k)} &\le \hat{J}^{(k)} + 2\left\| \Omega^{(k)} - \hat{\Omega}^{(k)} \right\| / T^{(k)} \\ &\le c_2^{(k)} J^{(k-1)} \left(\epsilon^{(k-1)} T^{(k)}\right) + 4d^{(k)} R\delta J / T^{(k)}. \end{aligned} \tag{151}$$

If at each level the second term in Eq. (151) is small compared to the first term, then our previous analysis of the pulse sequence remains a good approximation, and we conclude that the pulse errors do not compromise the effectiveness of concatenated DD very much. However, the second term imposes a floor on (our upper bound on) the effective noise strength

$$\eta_{\rm DD}^{(k)} = J^{(k)} T^{(k)} \ge 4d^{(k)} R\delta J \ge 4R\delta J. \tag{152}$$

As the level k increases, $\eta_{\rm DD}^{(k)}$ falls as in Eq. (138) as long as it remains well above the floor, but reaches a plateau as the floor is approached. Such behavior was observed in the numerical simulations reported in [10]. This floor might be substantially suppressed by using Eulerian pulse sequences as in Sec. VI C.

A noteworthy property of Eq. (151) is that only the pulse errors at the top level appear explicitly on the right-hand side. The errors at lower levels are included implicitly, through their contributions to $J^{(k-1)}$ and $\epsilon^{(k-1)}$. Accordingly, Eq. (151) captures the idea that the cumulative effect of the errors in the R^k pulses is smaller than might have been naively expected, because errors that occur at lower levels in the pulse sequence become suppressed by the upper level pulses. This is an important feature of concatenated DD.

IX. BEYOND THE LOCAL-BATH ASSUMPTION

A key element of the noise model formulated in Sec. II is the local-bath assumption: at any given time, the noise Hamiltonians H_a and H_b associated with distinct circuit locations a and b act not only on disjoint sets of qubits but also on disjoint baths. This assumption is important because it allows us to ignore interactions among different circuit locations and thus assign an effective noise strength $\eta_{\rm DD}$ to each DD-protected gate individually. The local-bath assumption may be a reasonable approximation to noise in actual systems, at least in some cases, but it is not strictly satisfied; surely there are bath degrees of freedom that couple to multiple qubits, even while these qubits are participating in distinct gates. Can our analysis be extended to noise models that include correlations that arise because qubits participating in different gates at the same time couple to common bath variables?

Accuracy threshold theorems have been proved for Hamiltonian models of correlated noise in [31, 32, 33]. Perhaps similar methods can be applied to DD-protected circuits, but this seems to be a technically challenging problem which we leave for the future.

However, there is an easier problem that already arises when we consider just a single circuit location, and disregard how the noise at one location is correlated with the noise at another location. How is our analysis affected if the qubits at this location couple not just to a local bath comprised of nearby bath degrees of freedom but to a global bath that includes bath variables that are far away? Of course, our previous analysis still applies if we replace the norm $\|H_{B,a}\|$ of the local-bath Hamiltonian by the norm $\|H_B\|$ of the global-bath Hamiltonian in Eq. (8) and Eq. (10), but the trouble with this approach is that $\|H_B\|$ is a huge number that scales linearly with the volume of the bath, while an accuracy threshold criterion should be stated in terms of intensive quantities that are independent of the size of the system and bath.

On the other hand, we expect on physical grounds that the bath has a decomposition into local subsystems, and that the coupling of a given bath subsystem to a system qubit decays as the distance increases between the bath subsystem and the qubit; if in contrast each system qubit were coupled with constant strength to bath subsystems arbitrarily far away, the noise would be unacceptably strong and coherent manipulation of the system would be hopeless. Even though the local-bath assumption formulated in Sec II may not hold exactly, a sensible noise model should be *quasi-local* — qubits ought to interact only very weakly with bath subsystems that are far away. In this case, can we express the effective noise strength in terms of intensive quantities?

To be concrete, consider a noise model in which a single system qubit is immersed in a bath of N_b non-interacting spins in an external magnetic field. The noise Hamiltonian, assuming $H_S = 0$, is

$$H = H_B + H_{SB} = \sum_i H_{B,i} + \sum_i H_{SB,i}, \tag{153}$$

where

$$\begin{aligned} H_{B,i} &\equiv \mathbb{I}_S \otimes B_i^0, \\ H_{SB,i} &\equiv \sum_\alpha \sigma^\alpha \otimes B_i^\alpha. \end{aligned} \tag{154}$$

Here, the index $i = 1, \ldots, N_b$ labels the bath spins and $\{\sigma_\alpha,\ \alpha = 1,2,3\}$ are the Pauli operators acting on the system qubit. We may define the strengths of the individual terms as

$$\lambda_i \equiv \|H_{SB,i}\| \quad \text{and} \quad b_i \equiv \|H_{B,i}\| = \left\|B_i^0\right\|, \tag{155}$$

and the strength of the system-bath coupling can be characterized by

$$J \equiv \sum_i \lambda_i \geq \|H_{SB}\|\,; \tag{156}$$

we assume that the sum converges to a (small) finite value in the limit $N_b \to \infty$. On the other hand, the quantity

$$\beta \equiv \sum_i b_i \geq \|H_B\| \tag{157}$$

is not expected to remain bounded as $N_b \to \infty$.

Now consider how the bath parameters $\{b_i\}$ enter the Magnus expansion for a DD memory sequence or for a DD-protected gate applied to the system qubit. The Hamiltonian $H_M(t)$ is

$$H_M(t) = H_B + H'(t) = \sum_i H_{B,i} + H'(t), \tag{158}$$

where $H'(t) = 0$ or $\tilde{H}_{\rm err}(t) (= \tilde{H}_{SB}(t))$ as in Sec. V A, so that $\|H'(t)\| \leq J$. Furthermore, bath operators acting on different bath spins commute:

$$[B_i^0, B_j^0] = [B_i^0, B_j^\alpha] = [B_i^\alpha, B_j^\alpha] = 0,\ \forall i \neq j; \tag{159}$$

the only nonvanishing commutators of bath operators are $[B_i^0, B_i^\alpha]$ and $[B_i^\alpha, B_i^\gamma]$ (for any spin i).

The bath parameters $\{b_i\}$ do not contribute to $\Omega_1(T)$, so consider $\Omega_2(T)$. To estimate the integral in Eq. (87) (taking $\Gamma = 0$ so that $H_M(t) = \tilde{H}(t)$), we need an upper bound on the commutators. We observe that

$$\begin{aligned} \|[H_B, \tilde{H}_{\rm err}(s_2)]\| &= \|\sum_i [H_{B,i}, \tilde{H}_{SB,i}(s_2)]\| \\ &\leq \sum_i 2\|H_{B,i}\| \cdot \|\tilde{H}_{SB,i}(s_2)\| \leq 2bJ, \end{aligned} \tag{160}$$

where we have defined the single-spin bath parameter

$$b \equiv \max_i \|H_{B,i}\|\,. \tag{161}$$

We also observe that

$$\begin{aligned} \|[\tilde{H}_{\rm err}(s_1), \tilde{H}_{\rm err}(s_2)]\| \leq \|\sum_{i,j} [\tilde{H}_{SB,i}(s_1), \tilde{H}_{SB,j}(s_2)]\| \\ \leq \sum_{i,j} 2\lambda_i\lambda_j = 2J^2. \end{aligned} \tag{162}$$

Together, Eq. (160) and Eq. (162) imply

$$\|[H_M(s_1), H_M(s_2)\| \leq 4bJ + 2J^2, \tag{163}$$

and plugging Eq. (163) into Eq. (87) yields

$$\|\Omega_2(T)\| \leq \left(4bJ + 2J^2\right) \frac{1}{4} T^2 \leq (JT)[(b+J)T]. \tag{164}$$

This result matches the conclusion we reached under the local-bath assumption (see Eq. (88) and Eq. (89)), but with β now replaced by b.

Similarly, upper bounds on the higher-order Magnus terms can be also be expressed in terms of J and b, though the "replace β by b rule" does not quite work beyond second order. Consider, for example, one triple commutator that occurs in $\Omega_3(T)$:

$$\begin{aligned} &\|[H_B, [\tilde{H}_{\rm err}(s_2), \tilde{H}_{\rm err}(s_3)]]\| \\ &= \|\sum_{ijk} [H_{B,k}, [\tilde{H}_{SB,i}(s_2), \tilde{H}_{SB,j}(s_3)]]\| \\ &\leq \sum_{ij} \Big(\|[H_{B,i}, [\tilde{H}_{SB,i}(s_2), \tilde{H}_{SB,j}(s_3)]]\| \\ &\qquad + \|[H_{B,j}, [\tilde{H}_{SB,i}(s_2), \tilde{H}_{SB,j}(s_3)]]\| \Big) \\ &\leq 2(2b) \sum_{i,j} 2\lambda_i\lambda_j = 8bJ^2. \end{aligned} \tag{165}$$

In contrast, in the local-bath model we could upper bound the corresponding triple commutator by $4\beta J^2$. Simply replacing β by b gives the wrong answer by a factor of 2, because it fails to take into account that there are two different bath spins that do not commute with $[\tilde{H}_{SB,i}(s_2), \tilde{H}_{SB,j}(s_3)]$ for $i \neq j$. Similar factors, dependent on n, occur in the higher-order nested commutators contributing to $\Omega_n(T)$, but these factors do not depend on the total number of bath spins N_b. We could also include quasi-local interactions among the bath spins, and still obtain an upper bound on each Magnus term expressed in terms of intensive quantities.

Even when our bounds on the Magnus expansion are intensive, they might still be useless, if each local bath subsystem has a Hamiltonian with a large norm. In that case, though, there is another method that might succeed, which relates the effective noise strength to the frequency spectrum of bath correlations. We turn to that method next.

X. DYNAMICAL DECOUPLING AND BATH CORRELATIONS

So far, we have described how to analyze the performance of DD using the toggling frame and the Magnus expansion. Another method is to use the interaction picture defined by $H_c(t) + H_B$; that is, to transform away both the control sequence acting on the system and the free bath dynamics. In that case, the interaction-picture Hamiltonian is

$$\begin{aligned}\tilde{H}(t) &= \left[U_c^\dagger(t) \otimes U_B^\dagger(t)\right] H_{\rm err} \left[U_c(t) \otimes U_B(t)\right] \\ &= \sum_\alpha S_\alpha(t) \otimes B_\alpha(t), \end{aligned} \tag{166}$$

where

$$S_\alpha(t) = U_c(t)^\dagger S_\alpha U_c(t), \quad B_\alpha(t) = e^{itB_0} B_\alpha e^{-itB_0}, \tag{167}$$

and we can study the interaction-picture time evolution operator using the Magnus expansion defined by this Hamiltonian. This expansion has the big advantage that the interaction picture sums up the effects of the free bath dynamics to all orders in β; therefore, higher-order corrections are small provided J is small, even though β may be large. But there is also a substantial disadvantage: because the interaction picture bath operator $B_\alpha(t)$ is now time dependent, a pulse sequence that achieves first-order decoupling in the toggling frame may not achieve first-order decoupling in the interaction picture.

On the other hand, if the bath operator $B_\alpha(t)$ is in some sense slowly varying, then first-order decoupling might be satisfied to a good approximation. Though the rate of change of the operator $B_\alpha(t)$ is actually of order β, if the *state* of the bath has suitable properties, then the expectation value of $B_\alpha(t)$ in that state may vary slowly; then DD may work well because the typical *frequencies* of the bath are sufficiently small, even though β may be large.

When estimating $\eta_{\rm DD}$ using the Magnus expansion, we did not make any assumption about the state of the bath. The new estimates we derive in this Section depend on the bath's frequency spectrum and hence implicitly on the bath's state. In order to obtain a simple formula for $\eta_{\rm DD}$ we will impose a further limitation on the noise model that was not needed in the Magnus expansion analysis — we assume that the state of the bath is discarded at the end of each circuit location, and replaced by a fresh bath state at the beginning of the next location. Thus we will include the effects of the bath's memory in analyzing the effectiveness of the DD pulse sequence at each circuit location, but we assume that noise correlations between consecutive circuit locations can be neglected. We recognize the artificiality of this noise model, but we adopt it anyway because it allows us to derive an explicit expression for $\eta_{\rm DD}$. See Appendix J for further discussion.

A. Dyson expansion

In the toggling frame, it is convenient to analyze DD using the Magnus expansion because for a well chosen sequence of ideal pulses Ω_1 is a pure bath term, and the remaining noise acting on the system resides in the higher order terms. But if we use the interaction picture instead, so that first-order decoupling is not exact even for ideal pulses, it is simpler to estimate the effective noise strength $\eta_{\rm DD}$ using the Dyson expansion rather than the Magnus expansion. The interaction-picture time evolution operator $\tilde{U}(t) = \left[U_c^\dagger(t) \otimes U_B^\dagger(t)\right] U(t,0)$ is

$$\tilde{U}(t) = \mathcal{T} \exp\left(-i \int_0^t dt' \tilde{H}(t')\right) \tag{168}$$

where $\mathcal{T}$ denotes time ordering. For the local-bath noise model, augmented by the assumption that the bath is refreshed at the beginning at each circuit location, the arguments in Appendix J show that the noise strength $\bar{\eta}$ can be expressed as

$$\begin{aligned}\bar{\eta}^2 &= \max_{a,|\Psi\rangle} \left\langle \left(\tilde{U}^\dagger(T) - \mathbb{I}\right)\left(\tilde{U}(T) - \mathbb{I}\right)\right\rangle \\ &= \max_{a,|\Psi\rangle} \left\langle 2\mathbb{I} - \tilde{U}(T) - \tilde{U}^\dagger(T)\right\rangle; \end{aligned} \tag{169}$$

here T is the duration of the location, the expectation value $\langle\cdot\rangle$ is evaluated in the pure state $|\Psi\rangle \otimes |\Phi_a\rangle$ where $|\Phi_a\rangle$ is (a purification of) the initial state of the bath at the beginning of location a, and the maximum is with respect to all circuit locations and all system states. As is also shown in Appendix J,

$$\begin{aligned}\bar{\eta}^2 &\le \max \int_0^T dt_1 \int_0^T dt_2 \left\langle \tilde{H}(t_1)\tilde{H}(t_2)\right\rangle \\ &+ 2\left(e^{JT} - 1 - JT - \frac{1}{2}(JT)^2\right). \end{aligned} \tag{170}$$

For each term in the expansion Eq. (166) the expectation value in the product state factorizes and we have

$$\begin{aligned}\bar{\eta}^2 \ &\leq \ \int_0^T dt_1 dt_2 \sum_{\alpha,\beta} \langle S_\alpha(t_1) S_\beta(t_2)\rangle_S \, \langle B_\alpha(t_1) B_\beta(t_2)\rangle_B \\ &+ \ 2\left(e^{JT} - 1 - JT - \frac{1}{2}(JT)^2\right), \end{aligned} \tag{171}$$

where the maximum over circuit locations and system states is implicit.

Now suppose that the bath's time correlations are stationary, *i.e.*, that the expectation value $\langle B_\alpha(t_1) B_\beta(t_2)\rangle_B$ is a function of the time difference $t_1 - t_2$; this will be true if the initial state of the bath commutes with H_B, for example if the state is a mixture of energy eigenstates such as a thermal state. Then the bath correlation function may be expressed as

$$\langle B_\alpha(t_1) B_\beta(t_2)\rangle_B = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega(t_1-t_2)} \ K_{\alpha\beta}(\omega), \tag{172}$$

and Eq. (171) becomes

$$\bar{\eta}^2 \leq \max \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sum_{\alpha,\beta} \langle \tilde{S}_\alpha(\omega) \tilde{S}_\beta(-\omega)\rangle_S \ K_{\alpha\beta}(\omega) + \ldots, \tag{173}$$

where

$$\tilde{S}_\alpha(\omega) = \int_0^T dt \ e^{-i\omega t} \ S_\alpha(t), \tag{174}$$

and the ellipsis indicates the terms higher order in J. Defining the bath's spectral function $J^2_{\alpha\beta,i}$ by

$$K_{\alpha\beta}(\omega) = 2\pi \sum_i J^2_{\alpha\beta,i} \delta(\omega - \omega_i), \tag{175}$$

our expression for (the square of) the noise strength is

$$\bar{\eta}^2 \leq \max \sum_{i,\alpha,\beta} J^2_{\alpha\beta,i} \langle \tilde{S}_\alpha(\omega_i) \tilde{S}_\beta(-\omega_i)\rangle_S + \cdots . \tag{176}$$

Thus, speaking loosely, DD is effective if $\tilde{S}_\alpha(\omega)$ is suppressed when ω is a "typical frequency" where the bath spectral function has support. We use the symbol J^2 advisedly, because $\sqrt{J^2_{\alpha\beta,i}}$, like $J = \max \|H_{\rm err}\|$, scales linearly with the strength of the system-bath coupling.

The operator $\tilde{S}_\alpha(\omega)$ can be written as $T\bar{S}_\alpha(\omega T)$, where $\bar{S}_\alpha$ is dimensionless. Adapting our terminology to this correlation function analysis, let us say that a pulse sequence achieves nth-order decoupling if the first n terms in the Taylor expansion of $\bar{S}_\alpha(\omega T)$ vanish, so that

$$\tilde{S}_\alpha(\omega) = T\left(\bar{S}_{\alpha,n}(\omega T)^n + O[(\omega T)^{n+1}]\right). \tag{177}$$

Equivalently, the pulse sequence achieves nth-order decoupling provided

$$\int_0^T dt \ t^m S_\alpha(t) = 0 \tag{178}$$

for all α and for $m = 0, 1, 2, \ldots, n-1$. Denoting the norm of the operator $\bar{S}_{\alpha,n}$ by $C_{\alpha,n}$, we find that for a pulse sequence achieving nth-order decoupling, the noise strength is

$$\eta_{\rm DD} \leq \left[\sum_{i,\alpha,\beta} C_{\alpha,n} C_{\beta,n} J^2_{\alpha\beta,i} T^2 \left(\omega_i T\right)^{2n}\right]^{1/2} + \ldots, \tag{179}$$

where now the ellipsis includes corrections both higher order in ωT and higher order in the Dyson expansion. Therefore, ignoring the $O[(JT)^3]$ corrections higher order in the Dyson expansion, nth-order decoupling implies that DD suppresses the effective noise strength by n powers of ωT where ω is a characteristic bath frequency, rather than n powers of ϵT as in our previous analysis using the Magnus expansion.

B. Universal decoupling

To be concrete, consider the case of a single qubit with noise Hamiltonian

$$H = \mathbb{I} \otimes B_0 + \sum_{\alpha=x,y,z} \sigma_\alpha \otimes B_\alpha. \tag{180}$$

For a sequence of ideal zero-width Pauli operator pulses, the time-dependent system operator in the interaction-picture Hamiltonian becomes

$$\sigma_\alpha(t) = F_\alpha(t)\sigma_\alpha, \tag{181}$$

where $F_\alpha(t) = \pm 1$ ($+1$ if σ_α commutes with $U_c(t)$ and -1 if σ_α anticommutes with $U_c(t)$). For the universal decoupling sequence

$$U_c(t_{\rm DD}) = Z\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}, \tag{182}$$

these functions are

$$\begin{aligned} F_x &= (++--), \\ F_y &= (+-+-), \\ F_z &= (+--+); \end{aligned} \tag{183}$$

here, for example, $F_x = (++--)$ means that F_x has the value $+1$ in the intervals $[0, \tau_0]$ and $[\tau_0, 2\tau_0]$ and has the value -1 in the intervals $[2\tau_0, 3\tau_0]$ and $[3\tau_0, 4\tau_0]$. All three functions integrate to zero over the interval $[0, 4\tau_0]$ and hence achieve first-order decoupling. Evaluating the Fourier transform

$$\tilde{F}_\alpha(\omega) = \int_0^{4\tau_0} dt \ e^{-i\omega t} F_\alpha(t), \tag{184}$$

we find

$$
\begin{aligned}
(-i\omega)\tilde{F}_x(\omega) &= (x-1)(1+x-x^2-x^3) = -(x^2-1)^2 = 4e^{-2i\omega\tau_0}\sin^2(\omega\tau_0) = 4(\omega\tau_0)^2+\ldots,\\
(-i\omega)\tilde{F}_y(\omega) &= (x-1)(1-x+x^2-x^3) = -(x-1)(x^4-1)/(x+1)\\
&= 2e^{-2i\omega\tau_0}\tan(\omega\tau_0/2)\sin(2\omega\tau_0) = 2(\omega\tau_0)^2+\ldots,\\
(-i\omega)\tilde{F}_z(\omega) &= (x-1)(1-x-x^2+x^3) = (x-1)(x^2-1)^2/(x+1)\\
&= 4ie^{-2i\omega\tau_0}\tan(\omega\tau_0/2)\sin^2(\omega\tau_0) = 2i(\omega\tau_0)^3+\ldots,
\end{aligned}
\tag{185}
$$

where $x = e^{-i\omega\tau_0}$. The low-frequency suppression of $\tilde{F}_y(\omega)$ is stronger by a factor of 2 than the suppression of $F_x(\omega)$ because the period of $F_y(t)$ is shorter than the period of $F_x(t)$. The function $\tilde{F}_z(\omega)$ is suppressed by a further power of $\omega\tau_0$ because $F_z(t)$ is time-symmetric: $F_z(4\tau_0 - t) = F_z(t)$. Indeed, for any function $F(t)$ satisfying $F(T-t) = F(t)$, we have

$$
\begin{aligned}
\tilde{F}(\omega) &= \int_0^T dt\; e^{-i\omega t}\; F(t) = \int_0^T dt\; e^{-i\omega t}\; F(T-t)\\
&= \int_0^T dt\; e^{-i\omega(T-t)}\; F(t) = e^{-i\omega T}\tilde{F}(-\omega);
\end{aligned}
\tag{186}
$$

thus $\tilde{F}(\omega) = e^{-i\omega T/2}\tilde{F}_{\rm even}(\omega)$, where $\tilde{F}_{\rm even}(\omega)$ is an even function of ω, and $\tilde{F}(\omega) = O(\omega^2)$ if $\tilde{F}(0)$ vanishes.

The time-symmetric pulse sequence

$$
U_c(t_{\rm DD}) = \mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}\mathbb{I}X\mathbb{I}Z\mathbb{I}X\mathbb{I}, \tag{187}
$$

achieves second-order decoupling because all three functions obey $F(t) = F(T-t)$:

$$
\begin{aligned}
F_x &= (++----++),\\
F_y &= (+-+--+-+),\\
F_z &= (+--++--+).
\end{aligned}
\tag{188}
$$

Compared to the four-pulse sequence, the functions F_x and F_y are repeated twice, but with a sign flip, so the Fourier transform is suppressed by an additional factor of $1 - x^4 = 2ie^{-2i\omega\tau_0}\sin(2\omega\tau_0) \approx 4i(\omega\tau_0)$. The function F_z is repeated without the sign flip, so its Fourier transform is multiplied by $1 + x^4 = 2e^{-2i\omega\tau_0}\cos(2\omega\tau_0) \approx 2$. Therefore we have

$$
\begin{aligned}
\tilde{F}_x(\omega) &= = -16\tau_0(\omega\tau_0)^2+\ldots,\\
\tilde{F}_y(\omega) &= = -8\tau_0(\omega\tau_0)^2+\ldots,\\
\tilde{F}_z(\omega) &= = -4\tau_0(\omega\tau_0)^2+\ldots.
\end{aligned}
\tag{189}
$$

Again, different types of low-frequency Pauli noise are suppressed by different (constant) factors, with the heaviest suppression for phase (*i.e.*, σ_z) noise. By altering the pulse sequence, the stronger suppression could be applied to σ_x or σ_y noise instead.

C. Finite-width pulses

If the pulses are not ideal, then first-order decoupling will not be exact. For example if the pulses have nonzero width, then there is a contribution to $\bar{\eta}^2$ of the form

$$
\begin{aligned}
&\int (dt_1 dt_2)_{\rm PW} \sum_{\alpha,\beta} \langle S_\alpha(t_1)S_\beta(t_2)\rangle_S \langle B_\alpha(t_1)B_\beta(t_2)\rangle_B\\
&= \sum_{i,\alpha,\beta} J^2_{\alpha\beta,i}\langle \tilde{S}_{\alpha,\rm PW}(\omega_i)\tilde{S}_{\beta,\rm PW}(-\omega_i)\rangle_S;
\end{aligned}
\tag{190}
$$

here $\int(dt)_{\rm PW}$ denotes integration over the nonzero-width pulses, and

$$
\tilde{S}_{\alpha,\rm PW}(\omega) = \int (dt)_{\rm PW}\; e^{-i\omega t}\; S_\alpha(t). \tag{191}
$$

For a sequence of N pulses, each with duration δ, we expect $\tilde{S}_{\alpha,\rm PW}(\omega) \approx N\delta\|S_\alpha\|$ for $\omega t \ll 1$. Comparing with Eq. (177), we conclude that for a pulse sequence that achieves nth-order decoupling in the ideal case, pulse-width corrections are small provided

$$
N\delta/T \ll (\omega T)^n \tag{192}
$$

where ω is a typical bath frequency. This is similar to the criterion we found using the Magnus expansion, except with the frequency ω now replacing the operator norm ϵ.

For an Eulerian sequence with reproducible pulse errors, $\tilde{S}_\alpha(\omega)$ vanishes in the limit $\omega \to 0$ (by the same reasoning as in Sec. VI C); therefore first-order decoupling is exact. Furthermore $\tilde{S}_\alpha(\omega)$ is an even function of ω for any time-symmetric pulse sequence, and therefore a time-symmetric Eulerian sequence achieves second-order decoupling.

D. Gaussian noise

We have seen that, while in our previous analysis we required $\beta = \max\|H_B\|$ to be small compared to $1/\tau_0$ in order to get a useful estimate of $\eta_{\rm DD}$, the analysis based on bath correlation functions can provide a useful estimate even if β is large. However we still require that $J = \max\|H_{\rm err}\|$ is small to justify neglecting the higher-order corrections in the Dyson expansion in Eq. (170). In some cases it is possible to go further and express these higher-order corrections in terms of correlation functions as well, thereby obtaining an estimate that makes sense even if the system qubits are coupled to bath operators with large norm (*e.g.*, the quadrature amplitudes of a bath of harmonic oscillators).

Consider, for example, a single qubit coupled to bath operators whose correlators obey Gaussian statistics in the interaction picture: the interaction-picture Hamiltonian is

$$\tilde{H}(t) = \sum_\alpha \sigma_\alpha(t) \otimes B_\alpha(t), \tag{193}$$

where the expectation value of an odd number of bath operators vanishes, and the expectation of an even number of bath operators is

$$\begin{aligned} &\langle B(1)B(2)\cdots B(2n)\rangle \\ &= \sum_{\text{contractions}} K(i_1,i_2)K(i_3,i_4)\cdots K(i_{2n-1},i_{2n}). \end{aligned} \tag{194}$$

Here the sum is over the $(2n)!/2^n n!$ ways to divide the labels $1,2,\ldots 2n$ into n unordered pairs, and we use the shorthand $B(i) = B_{\alpha_i}(t_i)$, $K(i,j) = \langle B(i)B(j)\rangle_B$. Thus terms of odd order in the Dyson expansion for $\bar{\eta}^2$ vanish, and we may bound the $(2n)$-th order term as

$$\begin{aligned} &\left|\left\langle \frac{1}{(2n)!}\int_0^T dt_1\ldots dt_{2n}\mathcal{T}\left(\tilde{H}(t_1)\cdots\tilde{H}(t_{2n})\right)\right\rangle\right| \le \frac{1}{(2n)!}\int_0^T dt_1\ldots dt_{2n}\sum_{\alpha_1,\ldots\alpha_{2n}}\sum_{\text{contractions}}|K(i_1,i_2)\cdots K(i_{2n-1},i_{2n})| \\ &= \frac{1}{(2n)!}\sum_{\text{contractions}}(2K)^n = \frac{K^n}{n!}, \quad \text{where} \quad K = \frac{1}{2}\int_0^T dtds\sum_{\alpha,\beta}\left|\langle B_\alpha(t)B_\beta(s)\rangle_B\right|. \end{aligned} \tag{195}$$

To derive Eq. (195), we use Eq. (194) and $\|\sigma_\alpha(t)\| = 1$, and we note that the value of $|K(i,j)|$ does not depend on the time ordering of t_i and t_j. We conclude that, in the case of Gaussian noise, the sum of all corrections higher than quadratic order in the Dyson expansion can be bounded above by

$$\sum_{n=2}^{\infty} K^n/n! = e^K - 1 - K, \tag{196}$$

and that the quadratic term provides a good approximation to the effective noise strength for K sufficiently small.

Another approach to analyzing DD is to use the Dyson expansion and to also expand $B_\alpha(t)$ in powers of $B_0 t$, thus obtaining a double expansion in powers of JT and βT. In that case we might say that "nth-order decoupling" is achieved if, in the expression for the interaction-picture evolution operator $\tilde{U}(T)$, all terms of order T^m are pure-bath terms for $m = 1,2,\ldots,n$. For the case of a qubit subject to pure dephasing noise ($B_1 = B_2 = 0$), it is shown in [13, 39] that in this sense nth-order decoupling can be achieved by a sequence of X pulses with $n+1$ pulse intervals, where the pulses are nonuniformly spaced in time. For general qubit noise, nth-order decoupling can be achieved by a sequence of nonuniformly spaced X and Z pulses with altogether $(n+1)^2$ pulse intervals [59].

The corrections higher-order in T are not necessarily small unless both $\beta T \ll 1$ and $JT \ll 1$ are satisfied. However, the ideal pulse sequence constructed in [59] has the property

$$\int_0^T dt\ t^m F_\alpha(t) = 0, \quad m = 0,1,\ldots n-1, \quad \alpha = x,y,z. \tag{197}$$

(The sequence in [13, 39] has this property only for $\alpha = z$.) Therefore, even if βT is not small, we can use the correlation function analysis to show that same sequence also achieves nth-order decoupling in the sense of Eq. (177). Therefore DD works effectively if $\omega T \ll 1$, where ω is a typical bath frequency, provided that either $JT \ll 1$ or (in the case of Gaussian noise) $K \ll 1$. The same remark applies to pure dephasing noise for the pulse sequence in [13, 39].

E. Concatenated dynamical decoupling

Instead of using the Magnus expansion, we can analyze the performance of concatenated DD sequences using the Dyson expansion and bath correlation functions. As in Sec. X A, we will suppose that the higher-order terms in the Dyson expansion can be neglected, and will focus on the lowest-order term Eq. (176). The objective is to show that, by concatenating k times a pulse sequence that achieves first-order decoupling, k-th order decoupling can be achieved, in the sense that $\tilde{S}_\alpha(\omega) = O[(\omega T)^k]$.

To illustrate the idea, consider the simple pulse sequence that decouples pure-dephasing noise for a single qubit:

$$U_c(t_{\text{DD}}) = X\mathbb{I}X\mathbb{I}, \tag{198}$$

so that the "level-1" function multiplying σ_z in the interaction picture can be represented as

$$F_z^{(1)} = (+-). \tag{199}$$

When we concatenate the pulse sequence, $F_z^{(1)}$ is replaced by $F_z^{(2)}$, in which $F_z^{(1)}$ is repeated twice, but with a sign flip in the second repetition:

$$F_z^{(2)} = (+--+), \tag{200}$$

and for higher-level sequences we have

$$\begin{aligned} F_z^{(3)} &= (+--+-++-), \\ F_z^{(4)} &= (+--+-++--++-+--+), \end{aligned} \tag{201}$$

etc. Evaluating the Fourier transforms of these functions,

$$\tilde{F}_z^{(k)}(\omega) = \int_0^{2^k\tau_0} dt\ e^{-i\omega t} F_z^{(k)}(t), \tag{202}$$

we see that

$$\tilde{F}_z^{(1)}(\omega) = (-i\omega)^{-1}(x-1)(1-x) \tag{203}$$

and

$$\tilde{F}_z^{(k)}(\omega) = \left(1 - x^{2^{k-1}}\right) \tilde{F}_z^{(k-1)}(\omega), \tag{204}$$

where $x = e^{-i\omega\tau_0}$, and hence

$$\begin{aligned} \tilde{F}_z^{(n)}(\omega) &= (-i\omega)^{-1} x^{1/2} \left(x^{1/2} - x^{-1/2}\right) \\ &\quad \times \prod_{k=1}^{n} x^{2^{k-2}} \left(x^{-2^{k-2}} - x^{2^{k-2}}\right) \\ &= 2\omega^{-1} x^{1/2} \sin(\omega\tau_0/2) \\ &\quad \times \prod_{k=1}^{n} x^{2^{k-2}} \left(2i \sin\left(2^{k-2}\omega\tau_0\right)\right). \end{aligned} \tag{205}$$

The leading behavior of this function for small $\omega\tau_0$ is

$$\tilde{F}_z^{(n)}(\omega) = \tau_0 (i)^n 2^{n(n-1)/2} \left(\omega\tau_0\right)^n + \cdots, \tag{206}$$

and therefore Eq. (176) becomes

$$\eta_{\rm DD}^{(n)} \le 2^{n(n-1)/2} \left(\sum_i \left(J_{33,i}^2 \tau_0^2\right) \left(\omega_i\tau_0\right)^{2n} \right)^{1/2} + \cdots, \tag{207}$$

where we neglect corrections both higher order in the Dyson expansion and higher order in frequency. Naively, this expression for the effective noise strength $\eta_{\rm DD}$ is optimized by choosing the level of concatenation n to be the largest integer such that $2^{n-1}(\omega\tau_0) < 1$ where ω is a "typical" bath frequency. Note, however, that for $2^n(\omega\tau_0) \approx 1$ the higher-order corrections in $(\omega\tau_0)$ modify $\eta_{\rm DD}$ by an $O(1)$ multiplicative factor. Note also that $2^n\tau_0 = T^{(n)}$ is the duration of the level-n pulse sequence, and thus the optimal pulse sequence has duration comparable to a typical inverse frequency of the bath.

Other concatenated pulse sequences can be studied similarly. Consider for example the universal DD sequence. We have seen in Eq. (185) that this sequence suppresses noise asymmetrically (the best suppression for σ_z, the worst for σ_x), so we might choose to alter the sequence at higher levels to provide more balanced noise suppression. But if we do not do that, the functions $\tilde{F}_\alpha^{(k)}(\omega)$ can be specified by augmenting Eq. (185) with

$$\begin{aligned} \tilde{F}_x^{(k)}(\omega) &= \left(1 + x^{4^{k-1}} - \left(x^{4^{k-1}}\right)^2 - \left(x^{4^{k-1}}\right)^3\right) \tilde{F}_x^{(k-1)}(\omega) \\ &= \left(x^{4^{k-1}}\right)^{3/2} (4i) \cos\left(4^{k-1}(\omega\tau_0)/2\right) \sin\left(4^{k-1}(\omega\tau_0)\right) \tilde{F}_x^{(k-1)}(\omega), \\ \tilde{F}_y^{(k)}(\omega) &= \left(1 - x^{4^{k-1}} + \left(x^{4^{k-1}}\right)^2 - \left(x^{4^{k-1}}\right)^3\right) \tilde{F}_y^{(k-1)}(\omega) = \left(x^{4^{k-1}}\right)^{3/2} \frac{(i)\sin\left(2\cdot 4^{k-1}(\omega\tau_0)\right)}{\cos\left(4^{k-1}(\omega\tau_0)/2\right)} \tilde{F}_x^{(k-1)}(\omega), \\ \tilde{F}_z^{(k)}(\omega) &= \left(1 - x^{4^{k-1}} - \left(x^{4^{k-1}}\right)^2 + \left(x^{4^{k-1}}\right)^3\right) \tilde{F}_z^{(k-1)}(\omega) = \left(x^{4^{k-1}}\right)^{3/2} \frac{(-2)\sin^2\left(4^{k-1}(\omega\tau_0)\right)}{\cos\left(4^{k-1}(\omega\tau_0)/2\right)} \tilde{F}_x^{(k-1)}(\omega), \end{aligned} \tag{208}$$

The weakest suppression of low-frequency noise occurs for $\tilde{F}_x^{(k)}(\omega)$, where

$$\tilde{F}_x^{(k)}(\omega) = \left(4^k\left(i\omega\tau_0\right) + \cdots\right) \tilde{F}_x^{(k-1)}(\omega), \tag{209}$$

and hence

$$\tilde{F}_x^{(k)}(\omega) = \tau_0 \prod_{k=1}^{n} \left(4^k i\omega\tau_0\right) + \cdots$$
$$= \tau_0 (i)^n 4^{n(n+1)/2} \left(\omega\tau_0\right)^n + \cdots, \quad (210)$$

where we neglect the terms higher order in $\omega\tau_0$. From Eq. (176) we obtain the estimate of the noise strength

$$\eta_{\rm DD}^{(n)} \approx 4^{n(n+1)/2} \left(\sum_{i,\alpha,\beta} \left(J_{\alpha\beta,i}^2 \tau_0^2\right) \left(\omega_i\tau_0\right)^{2n} \right)^{1/2} + \cdots. \quad (211)$$

Noting that the universal DD sequence has length $R = 4$, we see that Eq. (211) resembles Eq. (138), but with the operator norm ϵ replaced by a bath frequency.

XI. CONCLUSIONS

We have derived upper bounds on the effective noise strength $\eta_{\rm DD}$ for DD-protected quantum gates, in terms of the parameters of a Hamiltonian noise model. From the upper bounds on the noise strength we can extract a noise suppression threshold condition, a sufficient condition for DD-protected gates to outperform unprotected gates. We can also derive an accuracy threshold condition; when the noise parameters obey this condition, scalable quantum computing is possible. Our results show that DD, and in particular concatenated DD, can improve the gate accuracy and overhead cost of fault-tolerant quantum computing.

We have used two different methods to quantify the accuracy of DD-protected gates. From the Magnus expansion in the toggling frame we can derive an expression for $\eta_{\rm DD}$ in terms of the operator norm of the noise Hamiltonian; an advantage of this method is that $\eta_{\rm DD}$ does not depend on the state of the bath. From the Dyson expansion in the interaction picture we can derive an expression for $\eta_{\rm DD}$ in terms of the frequency spectrum of bath correlations. While the bath frequency spectrum does depend on the state of the bath, the second method sometimes yields useful result when the first method fails, because the norm of the noise Hamiltonian is too large.

Our analysis of fault-tolerant circuits built from DD-protected gates applies only to Hamiltonian noise models satisfying suitable assumptions. For the Magnus expansion analysis we used the local-bath model; this allows us to study each DD-protected gate individually, ignoring noise correlations among distinct gates being executed in parallel at the same time. For the correlation function analysis we used an even more artificial model, in which the state of the bath is refreshed after each DD-protected gate. This assumption allows us to include non-Markovian effects during the DD pulse sequence at each protected gate, but to ignore these effects when the DD-protected gates are composed in a quantum circuit. It is clearly desirable to extend our analysis to models with more general noise correlations.

Here we have proposed to combine DD with fault-tolerant quantum computing straightforwardly, by replacing each gate in a fault-tolerant circuit by the corresponding DD-protected gate. Perhaps more sophisticated ways to combine DD with fault tolerance can be found, leading to further gains in efficiency and accuracy.

Acknowledgments

Research of HKN and JP is supported by NSF under Grant No. PHY-0803371. JP's research is also supported by DOE under Grant No. DE-FG03-92-ER40701, and by NSA/ARO under Grant No. W911NF-09-1-0442. DAL thanks the Institute for Quantum Information at Caltech, where this work was done, and acknowledges funding from the US Department of Defense, NSF PHY-803304, NSF PHY-802678, and NSF CCF-726439.

APPENDIX A: REVIEW OF THE MAGNUS EXPANSION

Here we will briefly review some properties of the Magnus expansion that are used in our arguments. For a more detailed discussion, see [44].

The foundation of the Magnus expansion is this theorem:

Theorem 1. *Suppose*

$$\frac{d}{dt} e^{\Omega(t)} = M(t) e^{\Omega(t)}. \quad (A1)$$

Then

$$\frac{d}{dt}\Omega(t) = \sum_{n=0}^{\infty} \frac{B_n}{n!}\ {\rm ad}_{\Omega(t)}^n [M(t)]. \quad (A2)$$

Here the $\{B_n\}$ *are the Bernoulli numbers defined by*

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n, \quad (A3)$$

and ${\rm ad}_B^n[A]$ *is defined by*

$${\rm ad}_B^n[A] \equiv [B, [B, [\cdots [B, [B, A]] \cdots]]] \quad (A4)$$

(${\rm ad}_B^0[A] = A$*, and* ${\rm ad}_B^n[A]$ *for* $n \geq 1$ *contains* n *nested commutators). The series converges provided* $\|\Omega(t)\| < \pi$*.*

Proof. To obtain a useful expression for $M(t) =$

$\left(\frac{d}{dt}e^{\Omega(t)}\right)e^{-\Omega(t)}$, we first evaluate

$$\begin{aligned}
&\frac{d}{d\lambda}\left[\frac{d}{dt}\left(e^{\lambda\Omega(t)}\right)e^{-\lambda\Omega(t)}\right] \\
&= \left[\frac{d}{dt}\left(\frac{d}{d\lambda}e^{\lambda\Omega}\right)e^{-\lambda\Omega(t)}\right] - \left(\frac{d}{dt}e^{\lambda\Omega(t)}\right)\Omega(t)e^{-\lambda\Omega(t)} \\
&= \left[\frac{d}{dt}\left(e^{\lambda\Omega(t)}\Omega(t)\right)e^{-\lambda\Omega(t)}\right] - \left(\frac{d}{dt}e^{\lambda\Omega(t)}\right)\Omega(t)e^{-\lambda\Omega(t)} \\
&= e^{\lambda\Omega(t)}\left(\frac{d}{dt}\Omega(t)\right)e^{-\lambda\Omega(t)} \\
&= \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\mathrm{ad}^n_{\Omega(t)}\left[\frac{d}{dt}\Omega(t)\right]. \qquad \text{(A5)}
\end{aligned}$$

In the last line we have used the identity

$$e^{\lambda B}Ae^{-\lambda B} = \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\ \mathrm{ad}^n_B[A], \qquad \text{(A6)}$$

which can be verified by differentiating both sides k times with respect to λ and then setting $\lambda = 0$. Expressing $M(t)$ as the integral of its derivative, we find

$$\begin{aligned}
M(t) &= \int_0^1 d\lambda\frac{d}{d\lambda}\left[\frac{d}{dt}\left(e^{\lambda\Omega(t)}\right)e^{-\lambda\Omega(t)}\right] \\
&= \int_0^1 d\lambda\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\mathrm{ad}^n_{\Omega(t)}\left[\frac{d}{dt}\Omega(t)\right] \\
&= \sum_{n=0}^{\infty}\frac{1}{(n+1)!}\mathrm{ad}^n_{\Omega(t)}\left[\frac{d}{dt}\Omega(t)\right]. \qquad \text{(A7)}
\end{aligned}$$

Thus we have shown that $M(t) = \mathcal{O}_{\mathrm{ad}_{\Omega(t)}}\left[\frac{d}{dt}\Omega(t)\right]$, where

$$\mathcal{O}_A = \sum_{n=0}^{\infty}\frac{1}{(n+1)!}A^n = \frac{e^A - 1}{A}, \qquad \text{(A8)}$$

which is inverted by

$$\mathcal{O}_A^{-1} = \left(\frac{e^A-1}{A}\right)^{-1} = \sum_{n=0}^{\infty}\frac{B_n}{n!}A^n. \qquad \text{(A9)}$$

Therefore,

$$\frac{d}{dt}\Omega(t) = \mathcal{O}^{-1}_{\mathrm{ad}_{\Omega(t)}}\left[M(t)\right], \qquad \text{(A10)}$$

from which Eq. (A2) follows.

Regarding the convergence of the expansion, we note that $\|\mathrm{ad}^n_B\| \le (2\|B\|)^n$, and that the series expansion of $x/(e^x-1)$ converges for $|x| < 2\pi$, because the nearest poles to the origin in the complex x-plane are at $x = \pm 2\pi i$. Therefore the expansion in Eq. (A2) converges for $\|2\Omega(t)\| < 2\pi$.

□

In the Magnus expansion, we express $\Omega(t) = \sum_{n=1}^{\infty}\Omega_n(t)$, where $\Omega_n(t)$ is nth-order in M. Using this expansion, Eq. (A2) becomes

$$\begin{aligned}
\frac{d}{dt}\Omega_1(t) &= M(t), \\
\frac{d}{dt}\Omega_n(t) &= \sum_{j=1}^{n-1}\frac{B_j}{j!}S_n^{(j)}(t), \quad n \ge 2, \qquad \text{(A11)}
\end{aligned}$$

where

$$S_n^{(j)}(t) = \sum_{i_1,i_2,\ldots,i_j}^{(n-1)}\mathrm{ad}_{\Omega_{i_1}(t)}\mathrm{ad}_{\Omega_{i_2}(t)}\cdots\mathrm{ad}_{\Omega_{i_j}(t)}\left[M(t)\right]; \qquad \text{(A12)}$$

here the sum is over nonnegative integers $\{i_1, i_2, \ldots, i_j\}$ satisfying $i_1 + i_2 + \cdots + i_j = n-1$. We see that

$$S_n^{(1)}(t) = [\Omega_{n-1}(t), M(t)], \qquad \text{(A13)}$$

and that $S_n^{(j)}$ for $j > 1$ can be expressed as

$$S_n^{(j)} = \sum_{m=1}^{n-j}\left[\Omega_m(t), S^{(j-1)}_{n-m}(t)\right], \quad 2 \le j \le n-1. \qquad \text{(A14)}$$

The relations Eq. (A11), (A13), (A14) provide an algorithm for generating the terms in the Magnus expansion recursively, and we use these recursion relations to derive our upper bounds on the higher-order terms.

APPENDIX B: EVEN MAGNUS TERMS VANISH FOR A TIME-SYMMETRIC HAMILTONIAN

Here we prove the fact that, if $H_M(t)$ is time-symmetric, all even Magnus terms vanish. This was previously known in the NMR literature, at least for the case of a piecewise constant Hamiltonian [61].

Lemma 2. *If $H_M(T-t) = H_M(t)$, then $\Omega_n(T) = 0$ for all even n.*

Proof. First we show that $\Omega(T)$ is an odd function in $A(t) = -iH_M(t)$ when $H_M(t)$ (or correspondingly $A(t)$) is time-symmetric about $T/2$. Defining $\Delta_N \equiv T/2N$ for N a positive integer, the evolution operator from $t = 0$ to $t = T$ can be written as

$$\begin{aligned}
U(T,0) &= \lim_{N\to\infty} e^{A(T)\Delta_N}e^{A(T-\Delta_N)\Delta_N}\cdots e^{A(\frac{T}{2}+\Delta_N)\Delta_N} \\
&\qquad \times e^{A(\frac{T}{2}-\Delta_N)\Delta_N}\cdots e^{A(\Delta_N)\Delta_N}e^{A(0)\Delta_N} \\
&= \lim_{N\to\infty} e^{A(0)\Delta_N}e^{A(\Delta_N)\Delta_N}\cdots e^{A(\frac{T}{2}-\Delta_N)\Delta_N} \\
&\qquad \times e^{A(\frac{T}{2}-\Delta_N)\Delta_N}\cdots e^{A(\Delta_N)\Delta_N}e^{A(0)\Delta_N}, \qquad \text{(B1)}
\end{aligned}$$

where in the second equality, we have used the time-symmetry $A(T-t) = A(t)$. Taking the adjoint of Eq. (B1), and noting that $A(t)^\dagger = -A(t)$, we find

$$U^\dagger(T,0) = \lim_{N\to\infty} e^{-A(0)\Delta_N} e^{-A(\Delta_N)\Delta_N} \cdots e^{-A(\frac{T}{2}-\Delta_N)\Delta_N} \times e^{-A(\frac{T}{2}-\Delta_N)\Delta_N} \cdots e^{-A(\Delta_N)\Delta_N} e^{-A(0)\Delta_N}. \tag{B2}$$

Thus $U^\dagger(T,0)$ has the same form as $U(T,0)$, except for the replacement $A(t) \to -A(t)$.

Since $U(T,0) = \exp(\Omega(T))$ and $U^\dagger(T,0) = \exp(-\Omega(T))$, we conclude that under the replacement $A(t) \to -A(t)$, $\Omega(T)$ transforms as $\Omega(T) \to -\Omega(T) + i2\pi\ell$, for some integer ℓ. In fact, since the integer ℓ cannot jump discontinuously when $A(t)$ is smoothly deformed, ℓ must be a constant independent of $A(t)$, and by taking the limit $A(t) \to 0$ we see that $\ell = 0$; thus $\Omega(T)$ changes sign under $A(t) \to -A(t)$, i.e., is an odd function of $A(t)$.

In general, $\Omega_n(T)$ is an integral of an expression containing n factors of $A(t)$. Thus, $\Omega_n(T)$ is invariant under the replacement $A(t) \to -A(t)$ for n even, and changes sign under this replacement for n odd. Since in the time-symmetric case $\Omega(T)$ changes sign under $A(t) \to -A(t)$, we conclude that $\Omega_n(T)$ vanishes for n even. □

APPENDIX C: ERROR ESTIMATE FOR TIME EVOLUTION

Here we prove:

Lemma 3. *Suppose that the time evolution operator $U(t)$ satisfies the differential equation*

$$\frac{d}{dt}U(t) = -iH(t)U(t) \tag{C1}$$

with the initial condition $U(t_0) = U_0$, while $\tilde{U}(t)$ satisfies

$$\frac{d}{dt}\tilde{U}(t) = -i\tilde{H}(t)\tilde{U}(t) \tag{C2}$$

with the same initial condition, where both $H(t)$ and $\tilde{H}(t)$ are Hermitian. Then

$$\|\tilde{U}(t) - U(t)\| \le \int_{t_0}^t ds \|\tilde{H}(s) - H(s)\|. \tag{C3}$$

Proof.

$$\begin{aligned}
&\|\tilde{U}(t) - U(t)\| = \|\tilde{U}(t)U(t)^{-1} - \mathbb{I}\| \\
=\ & \left\| \int_{t_0}^t ds \frac{d}{ds}\left(\tilde{U}(s)U(s)^{-1}\right) \right\| \\
=\ & \left\| -i\int_{t_0}^t ds\ \tilde{U}(s)\left(\tilde{H}(s) - H(s)\right)U(s)^{-1} \right\| \\
\le\ & \int_{t_0}^t ds\ \left\| \tilde{U}(s)\left(\tilde{H}(s) - H(s)\right)U(s)^{-1} \right\| \\
\le\ & \int_{t_0}^t ds\ \left\| \tilde{H}(s) - H(s) \right\|.
\end{aligned} \tag{C4}$$

□

In this paper, we use Lemma 3 in three ways. In one application, we consider the case where both Hamiltonians are time independent, and conclude that (compare with Eq. (42))

$$\|\tilde{U}(t) - U(t)\| \le (t - t_0)\|\tilde{H} - H\|. \tag{C5}$$

This inequality allows us to relate the effective noise strength $\eta_{\rm DD}$ achieved by dynamical decoupling to our bounds on the terms in the Magnus expansion.

In another application, we consider H to be $H_S + H_B$, where H_S governs the ideal system dynamics and H_B governs the bath dynamics, while the Hamiltonian for the noisy joint evolution of system and bath is $\tilde{H} = H + H_{SB}$, where H_{SB} is responsible for the noise. Then in the local-bath model, if $\|H_{SB}\| \le J$ and a gate is executed in time τ_0, Lemma 3 implies that the norm of the "bad" part of the gate is bounded above by $J\tau_0$. Thus we may estimate the effective noise strength in the absence of DD as $\eta = J\tau_0$, as in Eq. (53).

In the third application, we use Lemma 3 to estimate the error arising from pulses with nonzero width. We consider $\tilde{H}(t)$ to be the Hamiltonian describing the actual DD sequence with realistic pulses, and $H(t)$ to be the idealized evolution for zero-width pulses, where both Hamiltonians are expressed in the toggling frame determined by the ideal sequence. Suppose that there are R pulses, and that each realistic pulse has support in a time interval of width δ. Both $\tilde{H}(t)$ and $H(t)$ can be expressed as a sum of a bath Hamiltonian and an error Hamiltonian; the bath Hamiltonian cancels in the difference $\tilde{H}(t) - H(t)$, and we suppose that for both the realistic and ideal sequences the norm of the error Hamiltonian is bounded above by J during the pulses. Thus $\|\tilde{H}(t) - H(t)\| \le 2J$ during the pulses (a total duration of $R\delta$), while $\tilde{H}(t) = H(t)$ outside the pulses; thus Lemma 3 implies $\|\tilde{U}(T) - U(T)\| \le 2R\delta J$, as in Eq. (149).

APPENDIX D: BOUNDS FOR EVEN MAGNUS TERMS IN THE TIME-SYMMETRIC CASE

We want to generalize the argument used to compute the bound for $\Omega_2(T)$ in the case where $H_M(t)$ is time-symmetric except for $t \in \Delta$. To do this for higher-order terms requires a formula for the Magnus terms for which all the multiple time-integrals are explicit. Such a formula can be found in [62] (for $n \ge 2$):

$$\Omega_n(T) = \frac{1}{n}\int_0^T dt_1 \ldots \int_0^T dt_n L_n \times [[\ldots[A(t_1), A(t_2)], \ldots], A(t_n)] \tag{D1}$$

where

$$L_n \equiv \sum_{l=1}^{n-1} \frac{1}{l}(-1)^{l+1} \sum_{1\le j_1<\ldots<j_{n-l}<n} \prod_{m=1}^{n-l} \Theta(j_m, j_m+1). \tag{D2}$$

The L_n coefficients take care of the time-ordering and relabeling of the integration variables. For n even, following what we did in the $\Omega_2(T)$ case, we split up the n time-integrals into n different cases: (1) none of $t_i, i=1,\ldots n$ are in Δ, (2) exactly one of $t_i \in \Delta$, (3) exactly two of $t_i \in \Delta$, …, (n) exactly n of $t_i \in \Delta$. Case (1) is zero from the time symmetry of $H_M(t)$ for $t \notin \Delta$; the remaining cases we bound by first bounding the nested commutator and L_n, and then doing the time-integral.

The $(n-1)$-nested commutator can be bounded as

$$\begin{aligned}
&\|[[\ldots[A(t_1), A(t_2)], \ldots], A(t_n)]\| \\
&\le 2^{n-2}\,\|[A(t_1), A(t_2)]\|\,\|A(t_3)\|\ldots\|A(t_n)\| \\
&\le 2^{n-2}(4J\epsilon)\epsilon^{n-2} \\
&= 2^n J\epsilon^{n-1}.
\end{aligned} \tag{D3}$$

The 2^{n-2} factor in the first line comes from opening up $(n-2)$-nested commutators using submultiplicativity of the operator norm. The $(4J\epsilon)$ factor in the second line comes from the bound for $\|[A(t_1), A(t_2)]\|$ given in Eq. (88). The coefficient L_n can be bounded by ignoring the step function (i.e. ignoring the time-ordering, since we do not have the details of Δ anyway):

$$\begin{aligned}
|L_n| &\le \sum_{l=1}^{n-1} \frac{1}{l} \sum_{1\le j_1<j_2<\ldots<j_{n-l}<n} 1 \\
&= \sum_{l=1}^{n-1} \frac{1}{l}\binom{n-1}{n-l} = \sum_{l=1}^{n-1} \frac{1}{n-l}\binom{n-1}{l}.
\end{aligned} \tag{D4}$$

The binomial factor arises from counting the number of terms in the sum over j_i: we pick $n-l$ elements from the numbers 1 to $n-1$, and arranging them in ascending order gives a single choice of $(j_1, j_2, \ldots, j_{n-l})$ and hence a single term in the sum. The number of ways of choosing $n-l$ elements from $n-1$ distinct numbers is given by the binomial factor. To bound the remaining sum, consider

$$\begin{aligned}
\int_0^1 dx(1+x)^{n-1} &= \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{1}{n-l} x^{n-l}\Big|_{x=0}^{x=1} \\
&= \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{1}{n-l}.
\end{aligned} \tag{D5}$$

Therefore, we have that

$$\begin{aligned}
|L_n| &\le \int_0^1 dx(1+x)^{n-1} - \binom{n-1}{0}\frac{1}{n} \\
&= \frac{2}{n}\left(2^{n-1}-1\right).
\end{aligned} \tag{D6}$$

Putting these back in $\Omega_n(T)$ (n even) and doing the time-integrals, we find that

$$\begin{aligned}
&\|\Omega_n(T)\| \\
&\le \frac{2}{n^2}\left(2^{n-1}-1\right)\left(2^n J\epsilon^{n-1}\right)\Bigg[\binom{n}{1}\Delta(T-\Delta)^{n-1} \\
&\qquad + \binom{n}{2}\Delta^2(T-\Delta)^{n-2} + \ldots + \binom{n}{n}\Delta^n\Bigg] \\
&= \frac{2^{n+1}J\epsilon^{n-1}}{n^2}\left(2^{n-1}-1\right)\left[T^n - (T-\Delta)^n\right].
\end{aligned} \tag{D7}$$

In the first inequality above, the terms in the brackets are the $n-1$ cases for choosing the times $t_1, \ldots, t_n$, with at least one being in Δ.

One can check that Eq. (D7) agrees with Eq. (104) in the $n=2$ case and with Eq. (106) in the $n=4$ case. We also see that, for each n, $\Omega_n(T)$ is of order ΔT^{n-1}, and thus vanishes in the limit $\Delta \to 0$.

APPENDIX E: $S_n^{(j)}$ COEFFICIENTS

Here we derive bounds on the $S_n^{(j)}$ coefficients found in the recursive formulas (Eqs. (82a) – (82e)) for the Magnus terms.

Lemma 4. *For all $n \ge 2$, $1 \le j \le n-1$,*

$$\|S_n^{(j)}(t)\| \le f_n^{(j)} J\,(2\epsilon t)^{n-1}\,, \tag{E1}$$

where the coefficients are defined recursively:

$$f_1^{(0)} = 1, \quad f_n^{(0)} = 0, \quad n > 1, \tag{E2a}$$

$$f_n^{(j)} = 2\sum_{m=1}^{n-j}\sum_{p=0}^{m-1} \frac{|B_p|}{p!m} f_m^{(p)} f_{n-m}^{(j-1)}, \quad n \ge 2; \tag{E2b}$$

here the $\{B_p\}$ are the Bernoulli numbers, defined by

$$\frac{x}{e^x-1} = \sum_{p=0}^{\infty} \frac{B_p}{p!}\,x^p. \tag{E3}$$

Proof. We will prove the lemma by induction. We begin with the smallest case where $n=2, j=1$:

$$\begin{aligned}
\|S_2^{(1)}(t)\| &= \|[\Omega_1(t), -iH_M(t)]\| \\
&\le \int_0^t ds\,\|\,[H_M(s), H_M(t)]\,\|.
\end{aligned} \tag{E4}$$

The commutator can be bounded as in Eq. (88): $\|[H_M(s), H_M(t)]\| \le 4J\epsilon$. This thus gives $\|S_2^{(1)}(t)\| \le 4J\epsilon t$. Since $f_2^{(1)} = 2$, this can be rewritten as $\|S_2^{(1)}\| \le 4J\epsilon t = f_2^{(1)} J(2\epsilon t)$.

For a given $n \ge 3$, suppose that the lemma holds for all $S_m^{(p)}$ for $m < n, 1 \le p \le m-1$. There are three different types of $S_n^{(j)}$:

$$S_n^{(1)}(t) = [\Omega_{n-1}(t), -iH_M(t)]\,; \tag{E5a}$$

$$S_n^{(n-1)}(t) = \left[\Omega_1(t), S_{n-1}^{(n-2)}(t)\right]; \tag{E5b}$$

$$S_n^{(j)}(t) = \left[\Omega_1(t), S_{n-1}^{(j-1)}(t)\right] + \sum_{m=2}^{n-j}\left[\Omega_m(t), S_{n-m}^{(j-1)}(t)\right], \quad \text{for } 2 \le j \le n-2. \tag{E5c}$$

Note that the last case occurs only for $n \ge 4$. We will bound each case separately. First, for $S_n^{(1)}$,

$$\begin{aligned}\|S_n^{(1)}(t)\| &\le 2\|\Omega_{n-1}(t)\|\|H_M(t)\| \\ &\le 2\epsilon\sum_{p=1}^{n-2}\frac{|B_p|}{p!}\int_0^t ds\|S_{n-1}^{(p)}\| \\ &\le J(2\epsilon t)^{n-1}\sum_{p=1}^{n-2}\frac{|B_p|}{p!(n-1)}f_{n-1}^{(p)}.\end{aligned} \tag{E6}$$

Eq. (E2) becomes $f_n^{(1)} = 2\sum_{p=1}^{n-2}\frac{|B_p|}{p!(n-1)}f_{n-1}^{(p)}$ when $j=1$; therefore $\|S_n^{(1)}(t)\| \le f_n^{(1)}J(2\epsilon t)^{n-1}$.

Next we bound $S_n^{(n-1)}$:

$$\begin{aligned}\|S_n^{(n-1)}(t)\| &\le 2\|\Omega_1(t)\|\|S_{n-1}^{(n-2)}(t)\| \\ &\le f_{n-1}^{(n-2)}J(2\epsilon t)^{n-1}\end{aligned} \tag{E7}$$

Eq. (E2) becomes $f_n^{(n-1)} = 2f_{n-1}^{(n-2)}$ when $j = n-1$; therefore $\|S_n^{(n-1)}(t)\| \le f_n^{(n-1)}J(2\epsilon t)^{n-1}$.

Lastly, the $2 \le j \le n-2$ cases:

$$\begin{aligned}&\|S_n^{(j)}(t)\| \\ &\le 2\|\Omega_1(t)\|\|S_{n-1}^{(j-1)}(t)\| \\ &\quad + 2\sum_{m=2}^{n-j}\sum_{p=1}^{m-1}\frac{|B_p|}{p!}\left(\int_0^t ds\|S_m^{(p)}(t)\|\right)\|S_{n-m}^{(j-1)}(t)\| \\ &\le f_{n-1}^{(j-1)}J(2\epsilon t)^{n-1} \\ &\quad + J^2t(2\epsilon t)^{n-2}\left[2\sum_{m=2}^{n-j}\sum_{p=1}^{m-1}\frac{|B_p|}{p!m}f_m^{(p)}f_{n-m}^{(j-1)}\right].\end{aligned} \tag{E8}$$

The expression within the brackets in the last line looks like $f_n^{(j)}$ in Eq. (E2), except we need to add in the $m=1$ terms, as well as the $p=0$ terms. In fact,

$$\begin{aligned}&2\sum_{m=2}^{n-j}\sum_{p=1}^{m-1}\frac{|B_p|}{p!m}f_m^{(p)}f_{n-m}^{(j-1)} \\ &= f_n^{(j)} - 2\frac{|B_0|}{0!1}f_1^{(0)}f_{n-1}^{j-1} - 2\sum_{m=2}^{n-j}\frac{|B_0|}{0!m}f_m^{(0)}f_{n-m}^{(j-1)} \\ &= f_n^{(j)} - 2f_{n-1}^{(j-1)},\end{aligned} \tag{E9}$$

where in the last line, we have used the fact that $f_{m>1}^{(0)} = 0$. Putting this into $\|S_n^{(j)}(t)\|$, and using the fact that $J \le \epsilon$, we get

$$\begin{aligned}\|S_n^{(j)}(t)\| &\le f_{n-1}^{(j-1)}J(2\epsilon t)^{n-1} \\ &\quad + J(\epsilon t)(2\epsilon t)^{n-2}\left[f_n^{(j)} - 2f_{n-1}^{(j-1)}\right] \\ &\le f_n^{(j)}J(2\epsilon t)^{n-1}.\end{aligned} \tag{E10}$$

This completes the induction. □

APPENDIX F: f_n COEFFICIENTS

In [58], the $\{f_n\}$ were shown to be coefficients in the power series expansion of of $G^{-1}(y) = \sum_{n=1}^{\infty} f_n y^n$, the inverse function of

$$y = G(s) = \int_0^s dx\left[2 + \frac{x}{2}\left(1 - \cot\frac{x}{2}\right)\right]^{-1}. \tag{F1}$$

Here, we will provide a self-contained proof of the above claim. It suffices to show that the coefficients of G^{-1} can be written in the form Eq. (96), with $f_n^{(j)}$ defined via the recursion relations (E2).

First, we prove a lemma that applies to a general function $y(s)$:

Lemma 5. *Suppose the smooth function $y \equiv G(s)$ is monotonic on its domain and satisfies $y(0) = 0$. Then $G^{-1}(y)$ can be written as $\sum_{n=1}^{\infty} f_n y^n$ where*

$$f_n = \frac{1}{n!}\left[\left(g(s)\frac{d}{ds}\right)^{n-1}g(s)\right]\Bigg|_{s=0}, \tag{F2}$$

and $\frac{dy}{ds} = \frac{1}{g(s)}$.

Proof. Since y is monotonic on its domain, the inverse function $G^{-1}(y)(= s)$ exists and has derivatives

$$\frac{d^n}{dy^n}G^{-1}(y) = \left(g(s)\frac{d}{ds}\right)^n s = \left(g(s)\frac{d}{ds}\right)^{n-1}g(s), \tag{F3}$$

where in the first equality, we have used the chain rule of differentiation: $\frac{d}{dy} = \frac{ds}{dy}\frac{d}{ds} = \left(\frac{dy}{ds}\right)^{-1}\frac{d}{ds} = g(s)\frac{d}{ds}$. Since y is a smooth function on its domain, so is $g(s)$ and hence all derivatives of $G^{-1}(y)$ exist. We can then expand $G^{-1}(y)$ as a Taylor series about $y=0$ and write $G^{-1}(y) = \sum_{n=0}^{\infty} f_n y^n$ for some coefficients f_n. We see that $f_0 = 0$ since $G^{-1}(0) = 0$. For $n \ge 1$, the Taylor coefficients are given by

$$f_n = \frac{1}{n!}\frac{d^n}{dy^n}G^{-1}(y)\Bigg|_{y=0}, \tag{F4}$$

which, upon inserting Eq. (F3) and noting that $y(0) = 0$, immediately gives Eq. (F2). □

For our purposes, the function $y(s)$ is given in Eq. (F1), i.e. $y(s) = G(s)$ which is smooth and monotonic over the domain $s \in [-2\pi, 2\pi]$. It is also clear that $y(0) = 0$. Lemma 5 thus tells us that we can write $G^{-1}(y) = \sum_{n=1}^{\infty} f_n y^n$, where f_n is given in Eq. (F2) with

$$g(s) \equiv \left(\frac{dy}{ds}\right)^{-1} = 2 + \frac{s}{2}\left(1 - \cot\frac{s}{2}\right). \tag{F5}$$

Lemma 6. *The coefficients f_n in $G^{-1}(y) = \sum_{n=1}^{\infty} f_n y^n$ can be written in the form Eq.* (96)*, with $f_n^{(j)}$ defined according to Eq.* (E2)*.*

Proof. For $n = 1$, the index j in Eq. (96) can only take value 0, so f_1 can be written in the form Eq. (96) if we set $f_1^{(0)} = 1$. To handle the case $n \geq 2$, we use Eq. (E3) to expand $\cot(s/2)$ in terms of Bernoulli numbers, finding

$$\frac{s}{2}\cot\left(\frac{s}{2}\right) = B_0 + \left(B_1 + \frac{1}{2}\right)(is) + \sum_{j=2}^{\infty} \frac{B_j}{j!}(is)^j. \tag{F6}$$

Noting that $B_0 = 1$, $B_1 = -1/2$, $B_{2j+1} = 0$ for $j \geq 1$, $B_{4j} < 0$ for $j \geq 1$ and $B_{4j+2} > 0$ for $j \geq 0$, Eq. (F6) becomes

$$g(s) = 2 + \frac{s}{2}\left(1 - \cot\frac{s}{2}\right) = \sum_{j=0}^{\infty} \frac{|B_j|}{j!} s^j. \tag{F7}$$

Using this series expansion of $g(s)$, we can rewrite (F2) for $f_{n\geq 2}$ as:

$$f_n = \frac{1}{n2^{n-1}} \sum_{j=1}^{n-1} \frac{|B_j|}{j!} \frac{2^{n-1}}{(n-1)!} \left[\left(g\frac{d}{ds}\right)^{n-1} s^j\right]\Bigg|_{s=0}. \tag{F8}$$

We omit the $j = 0$ term in Eq. (F8) because the derivative of a constant vanishes, and the sum over j terminates at $j = n-1$ because higher-order terms vanish when we set $s = 0$. Thus $f_{n\geq 2}^{(0)} = 0$, and by comparing with Eq. (96) we define $f_n^{(j)}$ as

$$f_n^{(j)} = \frac{2^{n-1}}{(n-1)!} \left[\left(g\frac{d}{ds}\right)^{n-1} s^j\right]\Bigg|_{s=0}. \tag{F9}$$

Now we need to show that the $\{f_n^{(j)}\}$ obey the recursive relation (E2). Using our definition of $f_n^{(j)}$ from (F9), the right-hand side of Eq. (E2) can be rewritten as

$$\begin{aligned}
&2\sum_{m=1}^{n-j}\sum_{p=0}^{m-1} \frac{|B_p|}{p!m} f_m^{(p)} f_{n-m}^{(j-1)} \\
&= 2\sum_{m=1}^{n-j}\sum_{p=0}^{m-1} \frac{|B_p|}{p!m} \frac{2^{m-1}}{(m-1)!} \left[\left(g\frac{d}{ds}\right)^{m-1} s^p\right]\Bigg|_{s=0} \\
&\qquad \times \frac{2^{n-m-1}}{(n-m-1)!} \left[\left(g\frac{d}{ds}\right)^{n-m-1} s^{j-1}\right]\Bigg|_{s=0} \\
&= \frac{2^{n-1}}{(n-1)!} \sum_{m=1}^{n-j} \binom{n-1}{m} \left[\left(g\frac{d}{ds}\right)^{m-1} \sum_{p=0}^{m-1} \frac{|B_p|}{p!} s^p\right]\Bigg|_{s=0} \\
&\qquad \times \left[\left(g\frac{d}{ds}\right)^{n-m-1} s^{j-1}\right]\Bigg|_{s=0}
\end{aligned} \tag{F10}$$

The expression $\left(\sum_{p=0}^{m-1} \frac{|B_p|}{p!} s^p\right)$ is just $g(s)$ if we can extend the upper limit of the sum to infinity. We can indeed do this, because in the equation above, the expression is differentiated $m-1$ times and s is set to 0. Hence, higher-order terms in the power series expansion of $g(s)$ with $p \geq m$ do not contribute. Therefore,

$$\begin{aligned}
&2\sum_{m=1}^{n-j}\sum_{p=0}^{m-1} \frac{|B_p|}{p!m} f_m^{(p)} f_{n-m}^{(j-1)} \\
&= \frac{2^{n-1}}{(n-1)!} \sum_{m=1}^{n-j} \binom{n-1}{m} \left[\left(g\frac{d}{ds}\right)^{m-1} g\right]\Bigg|_{s=0} \\
&\qquad \times \left[\left(g\frac{d}{ds}\right)^{n-m-1} s^{j-1}\right]\Bigg|_{s=0} \\
&= \frac{2^{n-1}}{(n-1)!} \sum_{m=1}^{n-1} \binom{n-1}{m} \left[\left(g\frac{d}{ds}\right)^{m} s\right]\Bigg|_{s=0} \\
&\qquad \times \left[\left(g\frac{d}{ds}\right)^{n-1-m} s^{j-1}\right]\Bigg|_{s=0}
\end{aligned} \tag{F11}$$

Now, for any differential operator $\mathcal{D}$ satisfying the product rule, i.e. $\mathcal{D}(xy) = \mathcal{D}(x)y + x\mathcal{D}(y)$ (where x and y commute), $\mathcal{D}^n$ has the binomial expansion

$$\mathcal{D}^n(xy) = \sum_{m=0}^{n} \binom{n}{m} \left[\mathcal{D}^m(x)\right]\left[\mathcal{D}^{n-m}(y)\right]. \tag{F12}$$

Take $\mathcal{D} = g\frac{d}{ds}$, $x = s$ and $y = s^{j-1}$. Then (note that the $m = 0$ term is zero),

$$\begin{aligned}
\left(g\frac{d}{ds}\right)^{n-1} s^j\Bigg|_{s=0} &= \sum_{m=0}^{n-1} \binom{n-1}{m} \left[\left(g\frac{d}{ds}\right)^{m} s\right]\Bigg|_{s=0} \\
&\qquad \times \left[\left(g\frac{d}{ds}\right)^{n-1-m} s^{j-1}\right]\Bigg|_{s=0}.
\end{aligned} \tag{F13}$$

Putting this into (F11) gives exactly the expression for $f_n^{(j)}$ in (F9). $\square$

APPENDIX G: PURE-BATH TERM IN SECOND ORDER OF THE MAGNUS EXPANSION

Here we consider the case where the toggling-frame Hamiltonian has a decomposition $H(t) = H_B + H_{\rm err}(t)$ such that

$$H_B = B_0 \otimes \mathbb{I}, \quad H_{\rm err}(t) = \sum_\alpha B_\alpha \otimes S_\alpha(t), \tag{G1}$$

where $S_\alpha(t) = U_c^\dagger(t) S_\alpha U_c(t)$ and the operators $\{S_\alpha\}$ are a Hermitian basis for traceless operators acting on the system such that

$$\mathrm{tr}\,(S_\alpha S_\beta) = 0 \tag{G2}$$

for all $\alpha \neq \beta$. We further assume that each pulse either commutes or anticommutes with each S_α, so that

$$S_\alpha(t) = U_c^\dagger(t) S_\alpha U_c(t) = \pm S_\alpha, \tag{G3}$$

and hence

$$H_{\rm err}(t) = \sum_\alpha \zeta_\alpha(t) B_\alpha \otimes S_\alpha \tag{G4}$$

where $\zeta_\alpha(t) = \pm 1$. These assumptions are true, in particular, for an n-qubit system if each S_α and each pulse is a traceless n-qubit Pauli operator. We will show that under these assumptions the second-order term in the Magnus expansion

$$\Omega_2(T) = -\frac{1}{2}\int_0^T dt_1 \int_0^{t_1} dt_2\ [H(t_1), H(t_{,2})] \tag{G5}$$

contains no pure bath term; that is, $\mathrm{tr}_S(\Omega_2(T)) = 0$.

Because $[H_B, H_B] = 0$, it suffices to show that the system trace vanishes for $[H_B, H_{\rm err}(t_1)]$ and $[H_{\rm err}(t_1), H_{\rm err}(t_2)]$ for any $t_1, t_2 \in [0, T]$. First we observe that

$$\begin{aligned}[H_B, H_{\rm err}(t)] &= \sum_\alpha \zeta_\alpha(t)[B_0 \otimes \mathbb{I}, B_\alpha \otimes S_\alpha] \\ &= \sum_\alpha \zeta_\alpha(t)[B_0, B_\alpha] \otimes S_\alpha\end{aligned} \tag{G6}$$

has vanishing system trace. Next we note that if the product $S_\alpha S_\beta$ is traceless it can be expanded in the basis $\{S_\alpha\}$, so that

$$S_\alpha S_\beta = \delta_{\alpha\beta} T_\alpha + \sum_\gamma g_{\alpha\beta\gamma} S_\gamma \tag{G7}$$

(where T_α might have a nonvanishing trace). Therefore

$$\begin{aligned}H_{\rm err}(t_1)H_{\rm err}(t_2) &= \sum_\alpha \zeta_\alpha(t_1)\zeta_\alpha(t_2) B_\alpha B_\alpha \otimes T_\alpha + \cdots \\ H_{\rm err}(t_2)H_{\rm err}(t_1) &= \sum_\alpha \zeta_\alpha(t_2)\zeta_\alpha(t_1) B_\alpha B_\alpha \otimes T_\alpha + \cdots\end{aligned} \tag{G8}$$

where the ellipsis represents terms with vanishing system trace. Thus in the commutator $[H_{\rm err}(t_1), H_{\rm err}(t_2)]$ the terms proportional to T_α cancel, and what remains has vanishing system trace, as we wished to show.

APPENDIX H: NOISE PARAMETERS FOR CONCATENATED DYNAMICAL DECOUPLING

For the analysis of concatenated DD in Sec. VIII, we considered dividing the third-order term in the Magnus expansion Ω_3 into a pure bath term and a remainder. For that purpose we may use:

Lemma 7. *Suppose an operator $\mathcal{O}$ has a decomposition*

$$\mathcal{O} = \mathbb{I} \otimes B_0 + \sum_\alpha S_\alpha \otimes B_\alpha, \tag{H1}$$

where both terms are Hermitian and $\mathrm{tr}\,(S_\alpha) = 0$ *for each* α. *Then*

$$\|B_0\| \leq \|\mathcal{O}\|, \tag{H2}$$

and

$$\|\sum_\alpha S_\alpha \otimes B_\alpha\| \leq 2\|\mathcal{O}\|. \tag{H3}$$

Proof. To derive Eq. (H2), suppose that $|\psi\rangle$ is a normalized pure state such that $|\langle\psi|B_0|\psi\rangle| = \|B_0\|$, and consider the expectation value

$$\begin{aligned}&\langle\chi| \otimes \langle\psi| \left(\sum_\alpha S_\alpha \otimes B_\alpha\right) |\chi\rangle \otimes |\psi\rangle \\ &= \langle\chi| \left(\sum_\alpha S_\alpha \langle\psi|B_\alpha|\psi\rangle\right) |\chi\rangle,\end{aligned} \tag{H4}$$

where $|\chi\rangle$ is also a normalized pure state. The right-hand side of Eq. (H4) is the expectation value in the state $|\chi\rangle$ of a traceless Hermitian operator. Unless this operator is zero, the expectation value can be either positive or negative depending on how $|\chi\rangle$ is chosen. By choosing $|\chi\rangle$ so that the expectation value $\langle\sum_\alpha S_\alpha \otimes B_\alpha\rangle$ in the state $|\chi\rangle \otimes |\psi\rangle$ is either zero or has the same sign as $\langle\mathbb{I} \otimes B_0\rangle$, we have

$$|\langle\mathcal{O}\rangle| \geq |\langle\mathbb{I} \otimes B_0\rangle|, \tag{H5}$$

and Eq. (H2) follows. From the triangle inequality,

$$\|\sum_\alpha S_\alpha \otimes B_\alpha\| = \|\mathcal{O} - \mathbb{I} \otimes B_0\| \leq \|\mathcal{O}\| + \|\mathbb{I} \otimes B_0\| \leq 2\|\mathcal{O}\|, \tag{H6}$$

which proves Eq. (H3). □

The inequality Eq. (H3) is tight if we do not restrict the dimension of the system, but if the system is a qubit (two dimensional), it can be improved to

$$\|\sum_\alpha \sigma_\alpha \otimes B_\alpha\| \leq \|\mathcal{O}\|. \tag{H7}$$

For a qubit, there is an anti-unitary time-reversal operator $T : |\psi\rangle \to \sigma_y|\psi\rangle^*$ such that $T^\dagger \sigma_\alpha T = -\sigma_\alpha$. Suppose $|\psi\rangle$ is a normalized pure state such that $|\langle\psi| \sum_\alpha \sigma_\alpha \otimes$

$B_\alpha|\psi\rangle| = \|\sum_\alpha \sigma_\alpha \otimes B_\alpha\|$. By applying $T\otimes\mathbb{I}$ if necessary, we can choose $|\psi\rangle$ so that $\langle\sum_\alpha \sigma_\alpha \otimes B_\alpha\rangle$ and $\langle\mathbb{I}\otimes B_0\rangle$ have the same sign (unless $\langle\mathbb{I}\otimes B_0\rangle = 0$). Therefore

$$|\langle\mathcal{O}\rangle| \geq \left|\left\langle \sum_\alpha \sigma_\alpha \otimes B_\alpha \right\rangle\right|, \tag{H8}$$

and Eq. (H7) follows.

APPENDIX I: RELATING DISTANCE BETWEEN OPERATORS TO DISTANCE BETWEEN THEIR EXPONENTIALS

Here we prove:

Lemma 8.

$$\begin{aligned}\left\|e^A - e^B\right\| &\geq 2\|A-B\| \\ &-2\exp\left(\frac{1}{2}\|A+B\|\right)\sinh\left(\frac{1}{2}\|A-B\|\right).\end{aligned} \tag{I1}$$

Proof. Expanding the exponentials, we obtain

$$e^A - e^B = A - B + \sum_{n=2}^{\infty}\frac{1}{n!}\left[A^n - B^n\right], \tag{I2}$$

and therefore

$$\|e^A - e^B\| \geq \|A-B\| - \sum_{n=2}^{\infty}\frac{1}{n!}\|A^n - B^n\|. \tag{I3}$$

Defining

$$N = \frac{1}{2}(A+B), \quad M = \frac{1}{2}(A-B), \tag{I4}$$

we have

$$A^n - B^n = (N+M)^n - (N-M)^n, \tag{I5}$$

and when we apply the binomial expansion to $(N+M)^n - (N-M)^n$ the terms even order in M cancel. There are $2\binom{n}{m}$ terms of order m in M for m odd, each with an operator norm bounded above by $\|M\|^m\|N\|^{n-m}$; therefore

$$\begin{aligned}\|A^n - B^n\| &\leq 2\sum_{\text{odd } m}\binom{n}{m}\|M\|^m\|N\|^{n-m} \\ &= (\|N\|+\|M\|)^n - (\|N\|-\|M\|)^n.\end{aligned} \tag{I6}$$

Thus we find

$$\begin{aligned}\sum_{n=2}^{\infty}\frac{1}{n!}\|A^n - B^n\| &\leq \exp(\|N\|+\|M\|) \\ &- \exp(\|N\|-\|M\|) - 2\|M\| \\ &= \exp(\|N\|)\cdot 2\sinh(\|M\|) - 2\|M\|,\end{aligned} \tag{I7}$$

and substituting into Eq. (I3) yields Eq. (I1). □

If the norm of the sum $A+B$ is not too large, we can use Lemma 8 to show that A is close to B when e^A is close to e^B. For example, suppose that

$$\|A+B\| \leq \epsilon_+, \quad \|A-B\| \leq \epsilon_-. \tag{I8}$$

Then Lemma 8 implies

$$\|A-B\| \leq c(\epsilon_+,\epsilon_-)\left\|e^A - e^B\right\|, \tag{I9}$$

where

$$c(\epsilon_+,\epsilon_-) = \left(2 - e^{\epsilon_+/2}\frac{\sinh\epsilon_-/2}{\epsilon_-/2}\right)^{-1}. \tag{I10}$$

For example, if $\epsilon_+ = \epsilon_- = 0.3$, we find $c(\epsilon_+,\epsilon_-) = 1.20$.

APPENDIX J: BATH-STATE-DEPENDENT NOISE STRENGTH AND DYSON EXPANSION

In the local-bath model, the noisy operation applied at the circuit location a is a unitary transformation $\overline{G}_a$ acting jointly on the system and bath. We may express $\overline{G}_a$ as the sum of a "good" part $\mathcal{G}_a = G_a\otimes B_a$ (where G_a is the ideal gate), and a "bad" part $\mathcal{B}_a = \overline{G}_a - \mathcal{G}_a$.

The accuracy threshold theorem proved in [31, 33] establishes that quantum computing is scalable provided the noise strength $\bar{\eta}$ is smaller than a critical value η_0. For this purpose, the noise strength may be defined in the following way. Recall that we model a noisy preparation of a qubit as an ideal preparation followed by noisy Hamiltonian evolution for a prescribed period. Therefore, we may assume that the initial state of the system at the very beginning of a quantum computation is ideal, and that the initial state of the system and bath is a product state

$$|\Psi_S^0\rangle\langle\Psi_S^0|\otimes\rho_B^0. \tag{J1}$$

It is convenient to introduce a reference system R that purifies the initial state of the bath; then the initial state of system, bath, and reference system is a pure state

$$|\Phi_{SBR}^0\rangle = |\Psi_S^0\rangle\otimes|\Upsilon_{BR}^0\rangle, \tag{J2}$$

where

$$\rho_B^0 = \mathrm{tr}_R\Big(|\Upsilon_{BR}^0\rangle\langle\Upsilon_{BR}^0|\Big). \tag{J3}$$

Now consider a quantum circuit acting on the initial state $|\Phi_{SBR}^0\rangle$, and let $\mathcal{I}_r$ denote a set of r locations in the circuit. Let $U^{\text{bad}}(\mathcal{I}_r)$ denote the transformation that results if we place the noisy gate $\overline{G}_a$ at each location $a\notin\mathcal{I}_r$ and place the bad part $\mathcal{B}_a$ at each location $a\in\mathcal{I}_r$; $U^{\text{bad}}(\mathcal{I}_r)$ acts trivially on R. We may say that the noise strength is $\bar{\eta}$ if

$$\|U^{\text{bad}}(\mathcal{I}_r)|\Phi^0\rangle\| \leq \bar{\eta}^r \tag{J4}$$

for any set $\mathcal{I}_r$ of r locations [33].

Since each $\overline{G}_a$ is unitary and therefore has operator norm 1, the submultiplicative property of the norm implies

$$\|U^{\mathrm{bad}}(\mathcal{I}_r)|\Phi^0\rangle\| \le \|U^{\mathrm{bad}}(\mathcal{I}_r)\| \le \prod_{a\in\mathcal{I}_r} \|\mathcal{B}_a\|. \tag{J5}$$

Therefore, we may choose the noise strength to be

$$\bar{\eta} = \max_a \|\mathcal{B}_a\|, \tag{J6}$$

which does not depend at all on the initial state of the bath [31]. We used this definition for the analysis in Sec. IV-VIII, based on the Magnus expansion, of the effective noise strength achieved by dynamical decoupling.

However, for the analysis of the effective noise strength in Sec. X, based on bath correlation functions and the Dyson expansion, we use a different definition of $\bar{\eta}$ that *does* depend on the initial state of the bath. To state the new definition simply, it is convenient to put a further limitation on the noise model that was not needed in the Magnus expansion analysis — we assume that the state of the bath is discarded at the end of each circuit location, and replaced by a fresh bath state at the beginning of the next location.

We admit that this new more restricted noise model is even more artificial than the local-bath model we analyzed previously using the Magnus expansion. We adopted the local-bath model so that we would not need to consider spatial correlations in the noise at any given time, and could therefore study the efficacy of the DD pulse sequence for each circuit location individually. But the local-bath model admits non-Markovian effects, temporal noise correlations arising from the bath's "memory." The theorem in [31], with the threshold condition expressed in terms of the noise strength defined as in Eq. (J6), applies to non-Markovian noise (and can in fact be formulated in a more general way, so that the local-bath assumption is not really needed). Therefore, by using the Magnus expansion to estimate the norm $\|\mathcal{B}_a\|$ achieved by DD at each location, we derived a threshold condition applicable to a noise model with limited spatial correlations but unlimited temporal correlations.

The situation gets more complicated if our characterization of the noise depends on the state of the bath, which evolves during the computation. The theorem in [33] expresses the threshold condition in terms of bath-state-dependent correlation functions, under the assumption that the noise is Gaussian. But extending that theorem to provide an expression for the DD-protected effective noise strength in a fully non-Markovian setting is a technically daunting problem that we have not solved. We will settle instead for a rather perverse compromise, in which the effects of the bath's memory are included in our analysis of the DD pulse sequence at each circuit location, but are assumed to be negligible when we stitch the DD-protected gates together in a quantum circuit.

Under the assumption that the bath's state is refreshed between consecutive circuit locations, the noisy operation at location a is applied to a product state, where the initial state $\rho_{B,a}$ of the local bath for location a does not depend on the noisy operations applied at earlier circuit locations. Under this assumption, Eq. (J4) is satisfied if we define

$$\bar{\eta} = \max_{a,|\Psi\rangle} \|\mathcal{B}_a\left(|\Psi\rangle\otimes|\Phi_a\rangle\right)\|, \tag{J7}$$

where $|\Phi_a\rangle$ is a purification of $\rho_{B,a}$, and the maximum is over all circuit locations and over all pure states of the system. In terms of the interaction-picture operator applied at location a,

$$\tilde{U}_a = \mathcal{G}_a^\dagger\overline{G}_a = \mathcal{G}_a^\dagger\left(\mathcal{G}_a+\mathcal{B}_a\right) = \mathbb{I}_a + \mathcal{G}_a^\dagger\mathcal{B}_a, \tag{J8}$$

we may write $\bar{\eta}$ as

$$\bar{\eta} = \max_{a,|\Psi\rangle} \left\|\left(\tilde{U}_a - \mathbb{I}_a\right)\left(|\Psi\rangle\otimes|\Phi_a\rangle\right)\right\|, \tag{J9}$$

or equivalently

$$\begin{aligned}\bar{\eta}^2 &= \max_{a,|\Psi\rangle}\left\langle\left(\tilde{U}_a^\dagger - \mathbb{I}_a\right)\left(\tilde{U}_a-\mathbb{I}_a\right)\right\rangle,\\ &= \max_{a,|\Psi\rangle}\left\langle 2\mathbb{I}_a - \tilde{U}_a - \tilde{U}_a^\dagger\right\rangle,\end{aligned} \tag{J10}$$

where $\langle\cdot\rangle$ denotes the expectation value in the state $|\Psi\rangle\otimes|\Phi_a\rangle$. This is the formula used in Sec. X.

For a location with duration T, the interaction-picture time-evolution operator is given by Dyson's formula

$$\tilde{U}(T) = \mathcal{T}\exp\left(-i\int_0^T dt\ \tilde{H}(t)\right) \tag{J11}$$

where $\mathcal{T}$ denotes time-ordering and $\tilde{H}(t)$ is the interaction-picture Hamiltonian, which obeys $\|\tilde{H}(t)\| = \|H_{\mathrm{err}}\| \le J$. Expanding the exponential, we find

$$\tilde{U}(T) = \mathbb{I} + \sum_{n=1}^{\infty}\tilde{U}_n(T), \tag{J12}$$

where

$$\tilde{U}_n(T) = \frac{(-i)^n}{n!}\int_0^T dt_1\cdots dt_n\mathcal{T}\left(\tilde{H}(t_1)\cdots\tilde{H}(t_n)\right), \tag{J13}$$

and hence

$$\|\tilde{U}_n(T)\| \le \frac{1}{n!}T^n\|\tilde{H}(t_1)\cdots\tilde{H}(t_n)\| \le \frac{(JT)^n}{n!}. \tag{J14}$$

Similarly, $U^\dagger(T)$ has the expansion

$$\tilde{U}^\dagger(T) = \mathbb{I} + \sum_{n=1}^{\infty}\tilde{U}_n^\dagger(T), \tag{J15}$$

where

$$\tilde{U}_n^\dagger(T) = \frac{(i)^n}{n!}\int_0^T dt_1\cdots dt_n\mathcal{T}'\left(\tilde{H}(t_1)\cdots\tilde{H}(t_n)\right); \tag{J16}$$

here $\mathcal{T}'$ denotes reverse-time ordering, and again

$$\|\tilde{U}_n^\dagger(T)\| \le \frac{(JT)^n}{n!}. \tag{J17}$$

Noting that $\tilde{U}_1(T) + \tilde{U}_1^\dagger(T) = 0$, we find

$$\begin{aligned} &\left\langle 2\mathbb{I} - \tilde{U}(T) - \tilde{U}^\dagger(T) \right\rangle \\ &\le -\left\langle \tilde{U}_2(T) + \tilde{U}_2^\dagger(T) \right\rangle + \sum_{n=3}^{\infty} \|\tilde{U}_n(T) + \tilde{U}_n^\dagger(T)\| \\ &\le -\left\langle \tilde{U}_2(T) + \tilde{U}_2^\dagger(T) \right\rangle + 2\sum_{n=3}^{\infty} \frac{(JT)^n}{n!} \\ &\le -\left\langle \tilde{U}_2(T) + \tilde{U}_2^\dagger(T) \right\rangle + 2\left(e^{JT} - 1 - JT - \frac{1}{2}(JT)^2\right). \end{aligned} \tag{J18}$$

To evaluate the expectation value of $\tilde{U}_2(T) + \tilde{U}_2^\dagger(T)$, we observe that

$$\begin{aligned} &\mathcal{T}\left(\tilde{H}(t_1)\tilde{H}(t_2)\right) + \mathcal{T}'\left(\tilde{H}(t_1)\tilde{H}(t_2)\right) \\ &= \tilde{H}(t_1)\tilde{H}(t_2) + \tilde{H}(t_2)\tilde{H}(t_1), \end{aligned} \tag{J19}$$

so that

$$\begin{aligned} &\tilde{U}_2(T) + \tilde{U}_2^\dagger(T) \\ &= -\frac{1}{2}\int_0^T dt_1 dt_2 \left(\tilde{H}(t_1)\tilde{H}(t_2) + \tilde{H}(t_2)\tilde{H}(t_1)\right) \\ &= -\int_0^T dt_1 dt_2 \tilde{H}(t_1)\tilde{H}(t_2). \end{aligned} \tag{J20}$$

Finally we may express the noise strength as

$$\begin{aligned} \bar{\eta}^2 \; &= \; \max \int_0^\infty dt_1 dt_2 \left\langle \tilde{H}(t_1)\tilde{H}(t_2) \right\rangle \\ &+ \; 2\left(e^{JT} - 1 - JT - \frac{1}{2}(JT)^2\right), \end{aligned} \tag{J21}$$

as in Eq. (170).

---

[1] U. Haeberlen, *High Resolution NMR in Solids*, Advances in Magnetic Resonance Series, Supplement 1 (Academic Press, New York, 1976).
[2] L. Viola and S. Lloyd, Phys. Rev. A **58**, 2733 (1998).
[3] P. Zanardi, Phys. Lett. A **258**, 77 (1999).
[4] L.-M. Duan and G. Guo, Phys. Lett. A **261**, 139 (1999).
[5] D. Vitali and P. Tombesi, Phys. Rev. A **59**, 4178 (1999).
[6] L. Viola, E. Knill, and S. Lloyd, Phys. Rev. Lett. **82**, 2417 (1999).
[7] L. Viola and E. Knill, Phys. Rev. Lett. **90**, 037901 (2003).
[8] M. S. Byrd and D. A. Lidar, Phys. Rev. A **67**, 012324 (2003).
[9] K. Khodjasteh and D. A. Lidar, Phys. Rev. Lett. **95**, 180501 (2005).
[10] K. Khodjasteh and D. A. Lidar, Phys. Rev. A **75**, 062310 (2007).
[11] L. Viola and E. Knill, Phys. Rev. Lett. **94**, 060502 (2005).
[12] O. Kern and G. Alber, Phys. Rev. Lett. **95**, 250501 (2009).
[13] G. S. Uhrig, Phys. Rev. Lett. **98**, 100504 (2007).
[14] A. J. Berglund, arXiv.org:quant-ph/0010001.
[15] E. M. Fortunato, L. Viola, J. Hodges, G. Teklemariam, and D. G. Cory, New J. Phys. **4**, 5.1 (2002).
[16] E. Fraval, M. J. Sellars, and J. J. Longdell, Phys. Rev. Lett. **95**, 030506 (2009).
[17] J. Petta, A. Johnson, J. Taylor, E. Laird, A. Yacoby, M. D. Lukin, C. Marcus, M. Hanson, and A. C. Gossard, Science **309**, 2180 (2005).
[18] J. Morton, A. Tyryshkin, A. Ardavan, S. Benjamin, K. Porfyrakis, S. Lyon, and G. Briggs, Nature Phys. **2**, 40 (2006).
[19] J. J. L. Morton, A. M. Tyryshkin, R. M. Brown, S. Shankar, B. W. Lovett, A. Ardavan, T. Schenkel, E. E. Haller, J. W. Ager, and S. A. Lyon, Nature **455**, 1085 (2008).
[20] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Nature **458**, 996 (2009).
[21] H. Uys, M. J. Biercuk, and J. J. Bollinger, Phys. Rev. Lett. **103**, 040501 (2009).
[22] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Phys. Rev. A **79**, 062324 (2009).
[23] S. Damodarakurup, M. Lucamarini, G. D. Giuseppe, D. Vitali, and P. Tombesi, Phys. Rev. Lett. **103**, 040502 (2009).
[24] S. E. Beavan, E. Fraval, M. J. Sellars, and J. J. Longdell, Phys. Rev. A **80**, 032308 (2009).
[25] Y. Sagi, I. Almog, and N. Davidson, eprint arXiv:0905.0286.
[26] *Fault-tolerant quantum computation* (IEEE Computer Society Press, Los Alamitos, CA, 1996).
[27] D. Aharonov and M. Ben-Or, SIAM J. Comput. **38**, 1207 (2008).
[28] A. Kitaev, Russ. Math. Surveys **52**, 1191 (1997).
[29] E. Knill, R. Laflamme, and W. H. Zurek, Proc. R. Soc. London, Ser. A **454**, 365 (1998).
[30] B. M. Terhal and G. Burkard, Phys. Rev. A **71**, 012336 (2005).
[31] P. Aliferis, D. Gottesman, and J. Preskill, Quant. Inf. Comp. **6**, 97 (2006).
[32] D. Aharonov, A. Kitaev, and J. Preskill, Phys. Rev. Lett. **96**, 050504 (2006).
[33] H. K. Ng and J. Preskill, Phys. Rev. A **79**, 032318 (2009).
[34] L. Viola, S. Lloyd, and E. Knill, Phys. Rev. Lett. **83**, 4888 (1999).
[35] M. S. Byrd and D. A. Lidar, Phys. Rev. Lett. **89**, 047901

(2002).
[36] K. Khodjasteh and D. A. Lidar, Phys. Rev. A **68**, 022322 (2003), erratum: *ibid*, Phys. Rev. A **72**, 029905 (2005).
[37] N. Boulant, M. A. Pravia, E. M. Fortunato, T. F. Havel, and D. G. Cory, Quant. Inf. Proc. **1**, 135 (2002).
[38] R. Bhatia, *Matrix Analysis*, no. 169 in Graduate Texts in Mathematics (Springer-Verlag, New York, 1997).
[39] W. Yang and R.-B. Liu, Phys. Rev. Lett. **101**, 180403 (2008).
[40] B. Lee, W. M. Witzel, and S. D. Sarma, Phys. Rev. Lett. **100**, 160505 (2008).
[41] G. S. Uhrig, Phys. Rev. Lett. **102**, 120502 (2009).
[42] G. S. Uhrig and S. Pasini, arXiv:0906.3605.
[43] W. Magnus, Comm. Pure Appl. Math. **7**, 649 (1954).
[44] S. Blanes, F. Casas, J. A. Oteo, and J. Ros, Phys. Reports **470**, 151 (2009).
[45] P. C. Moan, J. A. Oteo, and J. Ros, J. Phys. A **32**, 5133 (1999).
[46] S. Pasini, T. Fischer, P. Karbach, and G. S. Uhrig, Phys. Rev. A **77**, 032315 (2008).
[47] P. Aliferis and A. W. Cross, Phys. Rev. Lett. **98**, 220502 (2007).
[48] D. P. DiVincenzo and P. Aliferis, Phys. Rev. Lett. **98**, 020501 (2007).
[49] P. Wocjan, M. Rötteler, D. Janzing, and T. Beth, Phys. Rev. A **65**, 042309 (2002).
[50] M. Rötteler and P. Wocjan, IEEE Trans. Inform. Theory **52**, 4171 (2006).
[51] T. Gullion, D. B. Baker, and M. S. Conradi, J. Magn. Res. **89**, 479 (1990).
[52] M. Lizak, T. Gullion, and M. Conradi, J. Magn. Res. **91**, 254260 (1991).
[53] K. Khodjasteh and L. Viola, Phys. Rev. Lett. **102**, 080501 (2009).
[54] K. Khodjasteh and L. Viola, Phys. Rev. A **80**, 032314 (2009).
[55] K. Khodjasteh, D. Lidar, and L. Viola, arXiv:0908.1526.
[56] S. Klarsfeld and J. A. Oteo, Phys. Rev. A **39**, 3270 (1989).
[57] S. Blanes, F. Casas, J. A. Oteo, and J. Ros, J. Phys. A **31**, 259 (1998).
[58] P. C. Moan, Ph.D. thesis, University of Cambridge (2002).
[59] J. R. West, B. H. Fong, and D. A. Lidar, arXiv:0908.4490.
[60] A result similar to Eq. (137) was found in [10]; in our notation, Eq. (51) of [10] reads $J^{(k)} = (\beta\tau_0)^k R^{k(k+2)/2} J$, which was obtained under the assumption $\beta > J$, and with ϵ approximated by β.
[61] C. H. Wang and J. D. Ramshaw, Phys. Rev. B **6**, 3253 (1972).
[62] R. Suárez and L. Sáenz, J. Math. Phys. **42**, 4582 (2001).
[63] C. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).
[64] D. A. Lidar, P. Zanardi, and K. Khodjasteh, Phys. Rev. A **78**, 012308 (2008).