Uniformly Decaying Subspaces for Error Mitigated Quantum Computation

Nishchay Suri,[1,2,*] Jason Saied,[1] and Davide Venturelli[1,2]

[1]*QuAIL, NASA Ames Research Center, Moffett Field, California 94035, USA*

[2]*USRA Research Institute for Advanced Computer Science (RIACS), Mountain View, California 94043, USA*

(Dated: March 4, 2024)

We present a general condition to obtain subspaces that decay uniformly in a system governed by the Lindblad master equation and use them to perform error mitigated quantum computation. The expectation values of dynamics encoded in such subspaces are unbiased estimators of noise-free expectation values. In analogy to the decoherence free subspaces which are left invariant by the action of Lindblad operators, we show that the uniformly decaying subspaces are left invariant (up to orthogonal terms) by the action of the dissipative part of the Lindblad equation. We apply our theory to a system of qubits and qudits undergoing relaxation with varying decay rates and show that such subspaces can be used to eliminate bias up to first order variations in the decay rates without requiring full knowledge of noise. Since such a bias cannot be corrected through standard symmetry verification, our method can improve error mitigation in dual-rail qubits and given partial knowledge of noise, can perform better than probabilistic error cancellation.

Decoherence remains a major obstacle in performing useful quantum computation and thus keeping it from fulfilling its promises [1–9]. While significant efforts are being made to drive down the error rates at the hardware level, continued progress is taking place in the fields of quantum error correction (QEC) and quantum error mitigation (QEM) to combat realistic noise with the aim of using the current noisy intermediate scale quantum (NISQ) hardware to its full potential.

Powerful methods to protect quantum information were developed such as by encoding it in decoherence free subspaces (DFS) that are completely immune to noise [10–13]. Such subspaces arise as a result of symmetric noise usually resulting from collective decoherence mechanisms and therefore have limited applicability in addressing local noise affecting the NISQ hardware. In a similar spirit to DFS, symmetry verification protocols for error mitigation [14–16] are effective at identifying and removing specific errors making the system immune to noise that breaks the symmetry of the quantum state, but have no way of removing symmetry-preserving errors. In Ref. [17], the author develops a sampling based method allowing for partial cancellation of errors undetectable by symmetry verification. Error suppression methods such as dynamical decoupling (DD) [13, 18–20] can be effective at eliminating unwanted environmental interactions but are limited by their pulse widths, restricting its use given the high error rates in realistic noise. Widely used QEM methods such as zero noise extrapolation although simple to implement, require reliably scaling noise without changing its form and is limited in eliminating bias entirely. Finally, Probabilistic Error Cancellation (PEC) on the other hand, can completely eliminate bias, however demands full knowledge of noise that may be hard to attain [21, 22].

In this Letter, we show that for certain realistic local noise for which a DFS offering complete protection may not exist, we can obtain subspaces that are uniformly affected by noise. The expectation values of dynamics encoded in such *Uniformly Decaying Subspaces* (UDS) are unbiased estimators of noise-free expectation values and therefore can be used for error mitigation. Specifically, where DFS are left invariant by the action of Lindblad operators, we show that the uniformly decaying subspaces are left invariant (up to orthogonal terms) by the action of the dissipative part of the Lindblad equation. We present the general conditions for construction of such subspaces and apply our theory to a system of qubits and qudits undergoing strong individual relaxation with varying decay rates. The varying decay rates introduce a bias in estimation which cannot be detected through standard symmetry verification. We show that such a bias can be eliminated up to first order variations in decay rates using an ensemble of runs without requiring the knowledge of individual decay rates. Lastly we apply our theory to implement the transverse field Ising model (TFIM) in analog and circuit-model settings using dual-rail qubits. Our method improves on the standard symmetry verification for QEM used in dual-rail qubits and given partial knowledge of noise, can perform better than probabilistic error cancellation. We note that some earlier works have pointed out that coherences or off-diagonal parts of density matrix can decay with the same rate in *PT* symmetric Lindbladians [23, 24], with only specific observables with additional weak symmetry experiencing a uniform decay [25]. Our construction however does *not* have *PT* symmetry and ensures that the entire density matrix (off-diagonal and diagonal parts) and *all* observables experience a uniform decay.

Preliminaries- Consider a quantum system described by a density operator ρ interacting with the environment. Under the standard assumptions that a Markovian environment started to couple to the system in some past, the dynamics is completely positive and trace preserv-
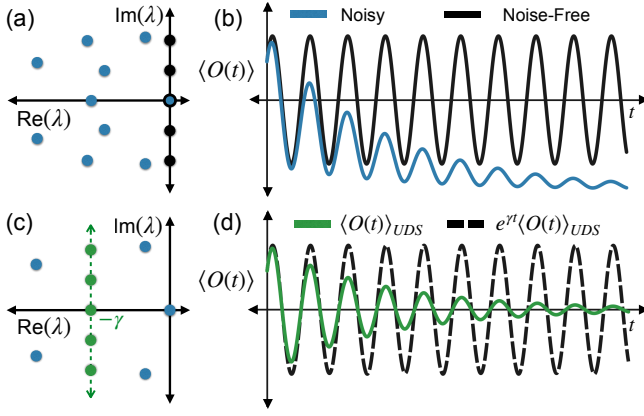
* nsuri@usra.edu

FIG. 1. (a) The distribution of eigenvalues of $\hat{\mathcal{L}}$ in the complex plane, with the black and blue circles indicating coherent and noisy dynamics and the corresponding (b) expectation values plotted with time. (c) The distribution of eigenvalues of $\hat{\mathcal{L}}$ for the uniformly decaying subspace marked by green circles and the corresponding (d) expectation value plotted with time.

ing, described by the Lindblad equation of the form

$$\dot{\rho} = \mathcal{L}\rho = \mathcal{H}\rho + \mathcal{D}\rho \; , \tag{1a}$$

$$\mathcal{D}\rho = \sum_{i=1}^{m} \gamma_i \mathcal{D}[A_i]\rho \; , \tag{1b}$$

where $\mathcal{H}\rho = -i[H,\rho]$ describes the unitarily evolving part under the Hamiltonian H and $\mathcal{D}[A]\rho = A\rho A^\dagger - \{A^\dagger A, \rho\}/2$ is the dissipative part with a general Lindblad operator A. We re-write the above equation in the Liouville space or third-quantization formalism [26–28] as $|\dot{\rho}\rangle\rangle = \hat{\mathcal{L}}|\rho\rangle\rangle$, where $\hat{\mathcal{L}} = \hat{\mathcal{H}} + \hat{\mathcal{D}}$. The expectation value of observable O is then given as

$$\langle O(t)\rangle = \langle\langle O|e^{\hat{\mathcal{L}}t}|\rho(0)\rangle\rangle = \sum_\alpha e^{\lambda_\alpha t}\langle\langle \ell_\alpha|\rho(0)\rangle\rangle\langle\langle O|r_\alpha\rangle\rangle , \tag{2}$$

where $|\ell_\alpha\rangle\rangle, |r_\alpha\rangle\rangle$ are bi-orthonormalized left and right eigenvectors of $\hat{\mathcal{L}}$ and λ_α is the corresponding eigenvalue.

Given the Hamiltonian H such that $H|E_\alpha\rangle = E_\alpha|E_\alpha\rangle$, the eigenvalues of the Liouville operator for coherent dynamics are purely imaginary, given by $\lambda^C_{\alpha\beta} = -i(E_\alpha - E_\beta)$ and marked by black circles in Fig. 1(a). The presence of noise causes the eigenvalues to be scattered in the complex plane as marked by the blue circles forcing different parts of the Hilbert space to decay at different rates and ruining the system's coherence. As an example, one qubit experiencing relaxation with decay rate γ already has two different decaying timescales $T_1 = 1/\gamma$ for diagonal and $T_2 = 1/(2\gamma)$ for off-diagonal parts of the density matrix. Coherence of systems is therefore preserved only for time scales $t \ll 1/\gamma$. The expectation value of an observable corresponding to the noisy and noise free cases are shown in Fig. 1(b).

In cases with high symmetry, usually resulting from collective decoherence, if the Lindblad operators A_i share a common set of eigenvectors $|\mathcal{P}_j\rangle$, then there exists

a DFS that is spanned by $\{|\mathcal{P}_j\rangle\}_{j=1}^N$ where $A_i|\mathcal{P}_j\rangle = c_i|\mathcal{P}_j\rangle, \forall i,j$. In general, it is sufficient to only require that the action of Lindblad operators leave a subsystem unaffected up to a gauge, such as $A_i|\mathcal{P}_j\rangle \otimes |g\rangle = c_i|\mathcal{P}_j\rangle \otimes |g'\rangle$, where $|g\rangle, |g'\rangle$ are vectors in the gauge space. This implies that the algebra generated by A_i leaves the subspace invariant, and the subspace is protected from decoherence as $\mathcal{D}|\mathcal{P}_i\rangle\langle\mathcal{P}_j| = 0, \forall i,j$. The Lindbladian is only composed of the Hamiltonian term and dynamics identical to noise-free case in Fig. 1 (a),(b).

Uniformly Decaying Subspaces- Analogous to the DFS, for certain special cases if instead the entire Lindblad dissipator operator $\mathcal{D}$ leaves a set of vectors $|\mathcal{P}_i\rangle\langle\mathcal{P}_j|$ unaffected up to an orthogonal component, then there exists a uniformly decaying subspace $\mathcal{S}$ that is spanned by $\{|\mathcal{P}_j\rangle\}_{j=1}^N$. The condition is given by

$$\mathcal{D}(|\mathcal{P}_i\rangle\langle\mathcal{P}_j| \oplus \mathcal{P}^\perp) = -\gamma|\mathcal{P}_i\rangle\langle\mathcal{P}_j| \oplus \mathcal{P}'^\perp , \tag{3}$$

$\forall i,j$ where $\mathcal{P}^\perp, \mathcal{P}'^\perp$ are in a subspace orthogonal to operators on $\mathcal{S}$. A sufficient condition can be obtained in terms of the Lindblad operators in a similar spirit to DFS. The computational subspace $\mathcal{S}$ is spanned by the eigenvectors of the summed operator $\sum_i A_i^\dagger A_i$ with the same eigenvalue: $\sum_i A_i^\dagger A_i|\mathcal{P}_j\rangle = -\gamma|\mathcal{P}_j\rangle$. Furthermore, there is a subspace $\mathcal{S}^\perp$, orthogonal to $\mathcal{S}$, with $A_i|\psi\rangle \in \mathcal{S}^\perp$ for any $|\psi\rangle \in \mathcal{S} \oplus \mathcal{S}^\perp$.

The condition Eq. 3 for uniformly decaying subspaces can be rewritten in the Liouville formalism as $\hat{\mathcal{P}}\hat{\mathcal{D}} = -\gamma\hat{\mathcal{P}}$, where $\hat{\mathcal{P}} = P \otimes P$ and $P = \sum_i |\mathcal{P}_i\rangle\langle\mathcal{P}_i|$ is the projector onto $\mathcal{S}$. We encode the initial density matrix and observable in the subspace: $|\rho(0)\rangle\rangle = \hat{\mathcal{P}}|\rho(0)\rangle\rangle$ and $|O\rangle\rangle = \hat{\mathcal{P}}|O\rangle\rangle$. Furthermore, we require that the Hamiltonian must commute with P, implying $\hat{\mathcal{P}}\hat{\mathcal{H}} = \hat{\mathcal{H}}\hat{\mathcal{P}} = \hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}$. Using the above, we evaluate the expectation value of the observable in the subspace after evolving for a time t [26]

$$\langle O(t)\rangle_{\mathrm{UDS}} = \langle\langle O|\hat{\mathcal{P}}e^{\hat{\mathcal{L}}t}\hat{\mathcal{P}}|\rho(0)\rangle\rangle , \tag{4a}$$

$$= \sum_n \frac{t^n}{n!}\langle\langle O|\hat{\mathcal{P}}(\hat{\mathcal{H}} - \gamma)^n\hat{\mathcal{P}}|\rho(0)\rangle\rangle , \tag{4b}$$

$$= e^{-\gamma t}\langle O\rangle_C , \tag{4c}$$

where $\langle O\rangle_C \equiv \langle\langle O|e^{\hat{\mathcal{H}}t}|\rho(0)\rangle\rangle$ is the noise-free expectation value. We therefore see that the expectation value of an observable encoded in the subspace is an unbiased estimator of noise-free expectation value up to an overall decaying factor. The eigenvalues of the encoded dynamics are $\lambda_{\alpha\beta} = -\gamma + i(E_\alpha - E_\beta)$ shown with green circles in Fig. 1 (c) and are identical those for coherent dynamics translated by a common constant. The measured expectation value in the subspace is shown in Fig. 1 (d) with the green line which when rescaled gives back the coherent dynamics shown with the dashed black line. The elimination of bias comes at the cost of requiring exponentially more measurements to obtain the result at same precision as time progresses, which is often true for error mitigation methods.

Rather than encoding the observable in the subspace or equivalently measuring POP directly, which can be difficult, for certain cases we can measure P, O simultaneously, post-selecting for the case $P = 1$. This follows the scheme behind symmetry verification and detecting erasure errors in dual-rail qubits [14–16]. The state is measured in the computational subspace $\mathcal{S}$ with probability $\langle\!\langle P|e^{\hat{\mathcal{L}}t}|\rho(0)\rangle\!\rangle = e^{-\gamma t}$, and the post-selected expectation value is $\langle O(t)\rangle_{\text{UDS}}^{\text{Post}} = \langle\!\langle O|\hat{\mathcal{P}}e^{\hat{\mathcal{L}}t}\hat{\mathcal{P}}|\rho(0)\rangle\!\rangle / \langle\!\langle P|e^{\hat{\mathcal{L}}t}|\rho(0)\rangle\!\rangle = \langle O\rangle_C$. The uniformly decaying part now represents the exponentially decreasing probability of finding the system in the correct subspace. Any errors taking the system out of computational subspace are removed by postselection. Under post-selection, the UDS conditions can be viewed as equivalent to that of symmetry verification with perfect error detection. We now extend its applicability by allowing for perturbations undetectable by symmetry verification.

Qubits/qudits under inhomogeneous relaxation- We now apply our theory to n qubits subject to strong individual relaxation, described by Eq. 1 with the dissipator operators $A_i = \sigma_i^- = |0\rangle\langle1|$ and decay rates given by γ_i. Such a noise model can be a generic realistic scenario in many NISQ devices. For example, dephasing rates were found to be an order of magnitude lower than relaxation rates in cQED superconducting qubits [29, 30] and photonic qubits in superconducting cavities where the dominant error is photon loss. Even in systems such as transmon qubits, where the dephasing rates are not too small as compared to relaxation rates, such a noise bias can be engineered by operating two transmons in resonance with each other forming the dual-rail qubits [31–33]. Moreover, DD can offer further protection from dephasing leading to an even greater increase in the bias [13, 18–20]. The decay rates can be different and even fluctuate over time [34].

We assume that we do not know the individual decay rates γ_i and only have the knowledge of the average decay rate $\bar{\gamma} = \frac{1}{N}\sum_{i=1}^{N}\gamma_i$. We parameterize $\gamma_i = \bar{\gamma} + \Delta\gamma_i$ and write Eq. 1 as $\mathcal{L}\rho = \mathcal{L}_0\rho + \mathcal{L}_1\rho$, where

$$\mathcal{L}_0\rho = \mathcal{H}\rho + \bar{\gamma}\sum_i \mathcal{D}[\sigma_i^-]\rho \,, \tag{5}$$

$$\mathcal{L}_1\rho = \sum_i \Delta\gamma_i \mathcal{D}[\sigma_i^-]\rho \,. \tag{6}$$

We provide the solution to the above equations by solving for the $\mathcal{L}_0\rho$ term exactly and treat the second term $\mathcal{L}_1\rho$ perturbatively. We require that the decay timescales of qubits are close to one another or $|\Delta\gamma|_{\max} \ll \bar{\gamma}, \|H\|$, where $\|\hat{\mathcal{L}}_1\| = |\Delta\gamma|_{\max} = \max_i |\Delta\gamma_i|$. The condition above poses no restrictions on the average decay time scales, which in principle can be large.

We consider the basis $|\mathcal{P}_{\boldsymbol{\alpha}}\rangle = |\alpha_1 \ldots \alpha_n\rangle$, where $\alpha_i \in \{0,1\}$ denotes the state of the ith qubit. We define our computational subspace $\mathcal{S}_k$ as the space spanned by all $|\mathcal{P}_{\boldsymbol{\alpha}}\rangle$ with $\sum_{i=1}^{n}\alpha_i = k$. In other words, each vector in this constant Hamming weight subspace has k qubits in state $|1\rangle$ and $n - k$ qubits in state $|0\rangle$. Therefore

the dimension of the computational subspace is given by $N = \dim(\mathcal{S}_k) \le \binom{n}{k}$. We see that the dissipation operator of the unperturbed term $\mathcal{D}_0 \equiv \bar{\gamma}\sum_i\mathcal{D}[\sigma_i^-]$ satisfies Eq. 3 with $\gamma = k\bar{\gamma}$ where $\mathcal{S}^\perp = \oplus_{i=0}^{k-1}\mathcal{S}_i$. Following the rest of the derivation, we see from Eq. 4 that $\langle O(t)\rangle_0 = \langle\!\langle O|\hat{\mathcal{P}}e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{P}}|\rho(0)\rangle\!\rangle = e^{-k\bar{\gamma}t}\langle O\rangle_C$. Therefore, given n qubits decaying with the same rate, we can obtain an unbiased estimation of expectation values by encoding the dynamics in $\mathcal{S}_k$. It is worthwhile to note that even for this apparently symmetric unperturbed term $\mathcal{D}_0$, there exists only a trivial DFS with one ket $|0\ldots0\rangle$, which cannot be used for computation.

We now consider the second term $\mathcal{L}_1$ perturbatively, calculating the bias in the estimation of expectation value of the observable $\langle O\rangle$ as $\Delta\langle O\rangle = \langle O\rangle_{\text{UDS}} - e^{-k\bar{\gamma}t}\langle O\rangle_C$. We have [26]

$$\Delta\langle O\rangle = \langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\int_0^t dt'\hat{\mathcal{L}}_1^I(t')|\rho(0)\rangle\!\rangle + \mathcal{O}(|\Delta\gamma|_{\max}^2)\,, \tag{7}$$

where $\hat{\mathcal{L}}_1^I(t) \equiv e^{-\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1 e^{\hat{\mathcal{L}}_0 t}$. We see that the variation of decay rates introduce additional bias that cannot be detected by just symmetry-verification. We define a unitary translation operator T that shifts the qubits by one position: $T|\alpha_1\alpha_2\ldots\alpha_{n-1}\alpha_n\rangle = |\alpha_n\alpha_1\ldots\alpha_{n-2}\alpha_{n-1}\rangle$. We use this shifted basis for computation, i.e. rotate the initial state, Hamiltonian and observable by T, we see that it leaves the unperturbed term invariant and only affects the perturbed term as

$$\mathcal{L}_1(T\rho T^\dagger) = T\sum_i \Delta\gamma_{i-1}\mathcal{D}[\sigma_i^-](\rho)T^\dagger \,. \tag{8a}$$

If we consider an ensemble of n evolutions with encodings shifted by one each corresponding to T^j where $j = 0, \ldots, n - 1$ and take the average of expectation values, the first order term is proportional to $\sum_{j=1}^{n}\Delta\gamma_{i-j} = \sum_{k=1}^{n}\Delta\gamma_k = 0$ as $\gamma_i = \bar{\gamma} + \Delta\gamma_i$. The bias $\Delta\langle O\rangle$ after taking the average over n evolutions is given by

$$e^{k\bar{\gamma}t}\langle\!\langle O\rangle\!\rangle_{\text{UDS}} - \langle O\rangle_C = \mathcal{O}\left(\frac{k^2|\Delta\gamma|_{\max}^2 t^2\|O\|}{2n}\right), \tag{9}$$

where $\langle\!\langle O\rangle\!\rangle_{\text{UDS}} = \frac{1}{n}\sum_{j=1}^{n}\langle O\rangle_{j,\text{UDS}}$. We therefore see that the bias is only second order in deviation of decay rates. The cancellation works similarly for the post selected case as well. Importantly, since the bias only depends on the deviation of the decay rates, an not on the magnitude of their average, it is possible to lower the bias arbitrarily by worsening the decay rates of all qubits until they align with the fastest decaying qubit, for example by tuning operating frequencies of the qubits. This comes at the expense of requiring more shots as the average decay rate increases. It is also noteworthy to point out that the knowledge of individual decay rates is not required to achieve this cancellation.

Furthermore, we also consider the more general case of amplitude damping channels where $A_i = a_i$ is the the bosonic annihilation operator. This noise model is especially important for photonic d-level qudits as photon
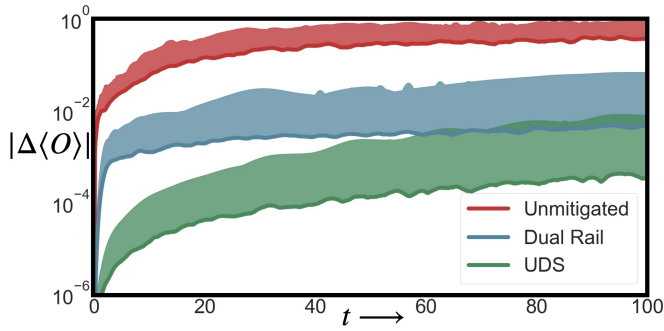
FIG. 2. We consider six physical qubits experiencing individual relaxation and encode a three qubit TFIM Hamiltonian in the dual-rail logical subspace. We chose initial state to be $|0\rangle_L^{\otimes 3}$ and measure the Z component of the first spin. The system is evolved using the Lindblad master equation and the modulus of the bias averaged over 100 random samples is shown by solid lines, with the boundary of shaded area indicating the worst case bias.

loss is the dominant source of noise [35, 36]. Our calculations generalize in a straightforward manner, with the details given in the supplemental material [26].

Examples- We now illustrate the QEM method by simulating the random transverse field Ising model (TFIM) in analog and circuit-model settings. Given n physical qubits, we choose $k = n/2$ to maximize the encoding efficiency, however we restrict our computational subspace to dual-rail qubits $\{|0\rangle_L \equiv |01\rangle, |1\rangle_L \equiv |10\rangle\}$ in order to efficiently implement the Ising Hamiltonian. The number of logical qubits is then limited to $n_L = n/2$ which is much less than $\binom{n}{n/2}$, but importantly this preserves the tensor product structure and allows the TFIM Hamiltonian to be realized using only two body local interactions. The TFIM Hamiltonian is given by

$$H = \frac{1}{2} \sum_{ij=1}^{n_L} J_{ij} Z_i^L Z_j^L + \frac{1}{2} \sum_{i=1}^{n_L} h_i^x X_i^L , \qquad (10)$$

where $Z_i^L \equiv Z_{2i}$ and $X_i^L \equiv (X_{2i} X_{2i+1} + Y_{2i} Y_{2i+1})/2$, are the logical operations on i^{th} qubit where $i = 0, \ldots, n_L - 1$. We can see that it preserves the dual-rail subspace and therefore commutes with the projector P to that subspace. We take 100 random instances with parameters J_{ij}, h_i^x chosen from a uniform distribution over $[-1, 1]$ and decay rates γ_i from a Gaussian distribution with mean 10^{-2} and standard deviation 10^{-3}.

Analog- We take six physical qubits that form three dual-rail logical qubits to simulate the TFIM Hamiltonian by solving the Lindblad equation. We choose the initial state to be $|0\rangle_L^{\otimes 3}$ and measure the Z component of the first spin. For the UDS and dual-rail method, we choose the observable $O = Z_0 \prod_{i=0}^{2} P_i$, where $P_i = \frac{I - Z_{2i} Z_{2i+1}}{2}$ takes values $\{1, 0\}$ dependent on whether the ith logical qubit is in the subspace. The observable is encoded in the computational subspace, and since it is a tensor product of Paulis, it can thus be measured efficiently. Equivalently,
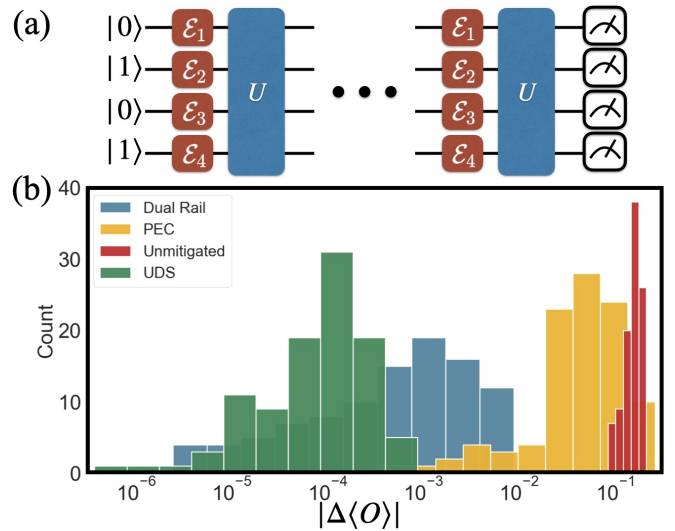

FIG. 3. We consider four physical qubits experiencing individual relaxation and encode a two qubit TFIM Hamiltonian in the dual-rail subspace. The Hamiltonian is implemented using 20 Trotter steps, where each unitary takes time 1/20 and is implemented as a product of one and two qubit gates. The amplitude damping channel acts on the ith qubit with damping rate γ_i. We choose the initial state $|0\rangle_L^{\otimes 2}$ and measure the qubits in computational basis. We allocate the same number of shots $N_{\text{shots}} = 10^7$ to each method and plot the histograms of the absolute bias of 100 random samples.

one may use post-selection by measuring all Z_i simultaneously, keeping only the measured value Z_0 when each $P_i = 1$. The modulus of the bias $|\Delta\langle O\rangle|$ is plotted in Fig. 2 on a log scale. The solid lines (bottom of the envelope) indicate the average bias of the 100 random samples and the boundary of the shaded region indicates the maximum bias. For reference, we plot the unmitigated bias which is calculated without any postselection. The dual-rail qubits use post-selection for symmetry verification eliminating any symmetry breaking errors. However due to the varying decay rates there is additional undetected error within the computational subspace. For the UDS method, we consider six separate evolutions with the encodings shifted by one each time as given in Eq. 8 and calculate their average as given by Eq. 9. This averaging procedure allows for the elimination of errors that are undetected by symmetry verification up to the first order and dramatically reduces the bias.

Circuit model- Where the previous example was focused on exact calculations of expectation values, we now show the applicability of our method on digital quantum computers with finite shots. We trotterize the evolution into N_T Trotter steps, applying N_T unitaries decomposed as one and two qubit gates with gate time t/N_T. Rather than evolving with respect to the Lindblad master equation Eq. 1, the dissipation is now modeled by the discrete amplitude damping channel $\mathcal{E}_i$ acting on the ith qubit with damping rate γ_i. We take four physical qubits, therefore encoding two dual-rail logical qubits,

and choose an illustrative case fixing $N_T = 20$ Trotter steps and time $t = 1$. We allocate the same number of shots $N_{\text{shots}} = 10^7$ [37] to each method and measure the Z component of the first spin same as above. We plot the histograms of the absolute bias of 100 random samples on a log scale in Fig. 3. We see that given the same number of shots, the UDS averaging method is able to perform better than dual-rail encoding alone by being able to eliminate first order bias undetected through symmetry verification. Since we also assume that only the average decay rates are known, probabilistic error cancellation can also only eliminate bias up to zeroth order. We can verify that the UDS results generalize to the discrete channel setting with the calculation details given in the supplementary [26].

Discussion- We presented conditions to obtain subspaces that decay uniformly and used them to perform error mitigated quantum computation. Where DFS are left invariant by the action of Lindblad operators, we show that the uniformly decaying subspaces are left invariant (up to orthogonal terms) by the action of the dissipative part of the Lindblad equation. For localized noise where a DFS offering complete protection may not exist, we show that we can still encode quantum information into UDS which can be used for error mitigation. We applied our theory to a system of qubits undergoing strong individual relaxation with varying decay rates and showed that the first order bias can be eliminated using an ensemble of runs without requiring the knowledge of individual decay rates. The subspaces therefore can be used to obtain unbiased estimators of the true expectation value up to second order variations in decay rates. We demonstrated our method by implementing the transverse field Ising model in analog and digital settings obtaining lower bias estimates as compared to dual-rail and probabilistic error cancellation methods. Our framework can be extended to other noise models, such as amplitude damping on qudits discussed in the supplementary material [26]. We also discuss its application to correlated decays such as $A_{ij} = \sigma_i^- \otimes \sigma_j^-$, which similarly give a UDS with cancellation of first-order errors. However, the cancellation becomes less straightforward with inefficient scaling. The techniques of UDS open up new possibilities in achieving lower biases in error mitigation. Since our method is a passive encoding scheme, i.e. it does not require active measurement-feedback, it can be used as a base encoding compatibly with other QEM or QEC methods.

[1] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).

[2] T. Pellizzari, S. A. Gardiner, J. I. Cirac, and P. Zoller, Decoherence, continuous observation, and quantum computing: A cavity qed model, Physical Review Letters **75**, 3788 (1995).

[3] I. L. Chuang, R. Laflamme, P. W. Shor, and W. H. Zurek, Quantum computers, factoring, and decoherence, Science **270**, 1633 (1995).

[4] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).

[5] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Applied physics reviews **6** (2019).

[6] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annual Review of Condensed Matter Physics **11**, 369 (2020).

[7] H.-P. Breuer and F. Petruccione, *The theory of open quantum systems* (Oxford University Press, USA, 2002).

[8] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Reviews of Modern Physics **82**, 1155 (2010).

[9] D. A. Lidar, Lecture notes on the theory of open quantum systems, arXiv preprint arXiv:1902.00967 (2019).

[10] D. A. Lidar, I. L. Chuang, and K. B. Whaley, Decoherence-free subspaces for quantum computation, Physical Review Letters **81**, 2594 (1998).

[11] S. Wu, L. Wang, and X. Yi, Time-dependent decoherence-free subspace, Journal of Physics A: Mathematical and Theoretical **45**, 405305 (2012).

[12] P. G. Kwiat, A. J. Berglund, J. B. Altepeter, and A. G. White, Experimental verification of decoherence-free subspaces, Science **290**, 498 (2000).

[13] D. A. Lidar, Review of decoherence-free subspaces, noiseless subsystems, and dynamical decoupling, Quantum information and computation for chemistry , 295 (2014).

[14] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Low-cost error mitigation by symmetry verification, Phys. Rev. A **98**, 062339 (2018).

[15] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. De Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Physical Review A **95**, 042308 (2017).

[16] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Phys. Rev. Lett. **122**, 180501 (2019).

[17] Z. Cai, Quantum error mitigation using symmetry expansion, Quantum **5**, 548 (2021).

[18] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling

of open quantum systems, Physical Review Letters **82**, 2417 (1999).

[19] G. De Lange, Z.-H. Wang, D. Riste, V. Dobrovitski, and R. Hanson, Universal dynamical decoupling of a single solid-state spin from a spin bath, Science **330**, 60 (2010).

[20] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Optimized dynamical decoupling in a model quantum memory, Nature **458**, 996 (2009).

[21] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Physical review letters **119**, 180509 (2017).

[22] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, arXiv:2210.00921 (2022).

[23] T. Prosen, P t-symmetric quantum liouvillean dynamics, Physical Review Letters **109**, 090404 (2012).

[24] T. Prosen, Generic examples of p t-symmetric qubit (spin-1/2) liouvillian dynamics, Physical Review A **86**, 044103 (2012).

[25] M. van Caspel and V. Gritsev, Symmetry-protected coherent relaxation of open quantum systems, Physical Review A **97**, 052106 (2018).

[26] Supplmenetary Materials.

[27] M. Krishna, P. Solanki, and S. Vinjanampathy, Select topics in open quantum systems, Journal of the Indian Institute of Science , 1 (2022).

[28] A. C. Li, F. Petruccione, and J. Koch, Perturbative approach to markovian open quantum systems, Scientific reports **4**, 4887 (2014).

[29] F. Yan, D. Campbell, P. Krantz, M. Kjaergaard, D. Kim, J. L. Yoder, D. Hover, A. Sears, A. J. Kerman, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Distinguishing coherent and thermal photon noise in a circuit quantum electrodynamical system, Phys. Rev. Lett. **120**, 260504 (2018).

[30] J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, Decoherence benchmarking of superconducting qubits, npj Quantum Information **5**, 54 (2019).

[31] A. Kubica, A. Haim, Y. Vaknin, F. Brandao, and A. Retzker, Erasure qubits: Overcoming the t_1 limit in superconducting circuits, arXiv preprint arXiv:2208.05461 (2022).

[32] D. L. Campbell, Y.-P. Shim, B. Kannan, R. Winik, D. K. Kim, A. Melville, B. M. Niedzielski, J. L. Yoder, C. Tahan, S. Gustavsson, *et al.*, Universal nonadiabatic control of small-gap superconducting qubits, Physical Review X **10**, 041051 (2020).

[33] H. Levine, A. Haim, J. S. Hung, N. Alidoust, M. Kalaee, L. DeLorenzo, E. A. Wollack, P. A. Arriola, A. Khalajhedayati, Y. Vaknin, *et al.*, Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons, arXiv:2307.08737 (2023).

[34] M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, Dynamics of superconducting qubit relaxation times, npj Quantum Information **8**, 132 (2022).

[35] I. L. Chuang, D. W. Leung, and Y. Yamamoto, Bosonic quantum codes for amplitude damping, Physical Review A **56**, 1114 (1997).

[36] C. Gardiner and P. Zoller, *Quantum noise: a handbook of Markovian and non-Markovian quantum stochastic methods with applications to quantum optics* (Springer Science & Business Media, 2004).

[37] The results remain qualitatively the same for much less number of shots, we chose 10^7 to clearly show the plots on the log scale.

Supplemental Material for "Uniformly Decaying Subspaces for Error Mitigated Quantum Computation"

Nishchay Suri,[1,2,*] Jason Saied,[1] and Davide Venturelli[1,2]

[1] QuAIL, NASA Ames Research Center, Moffett Field, California 94035, USA

[2] USRA Research Institute for Advanced Computer Science, Mountain View, California 94043, USA

(Dated: March 4, 2024)

I. LIOUVILLE SPACE REPRESENTATION

Let $\mathcal{W}$ be a Hilbert space of pure states. For any subspace $\mathcal{V} \subseteq \mathcal{W}$, let $\mathscr{L}(\mathcal{V}) = \mathrm{End}(V)$ denote the space of linear operators $\mathcal{V} \to \mathcal{V}$.

For $\rho \in \mathscr{L}(\mathcal{W})$, the Lindblad master equation is given as $\dot{\rho} = \mathcal{L}\rho = \mathcal{H}\rho + \mathcal{D}\rho$, such that

$$\mathcal{H}\rho = -i[H, \rho] \,, \tag{S1}$$

$$\mathcal{D}\rho = \sum_i \gamma_i \mathcal{D}[A_i]\rho \,, \tag{S2}$$

where $\mathcal{D}[A]\rho = A\rho A^\dagger - \frac{1}{2}\{A^\dagger A, \rho\}$. In many scenarios, analyzing the equation and its evolution becomes easier in the vectorized or Liouville space representation also known as the third-quantization. The rows of $N \times N$ density matrix ρ can be stacked on top of one another to obtain its vectorized form $|\rho\rangle\rangle$ with dimension $N^2 \times 1$, identifying $\mathscr{L}(\mathcal{V})$ with $V \otimes V^*$. As an example the 2×2 density matrix transforms as $\rho = \begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix} \iff |\rho\rangle\rangle = \begin{bmatrix} \rho_{00} & \rho_{01} & \rho_{10} & \rho_{11} \end{bmatrix}^T$. Given a superoperator $\mathcal{F}$ on $\mathscr{L}(\mathcal{L})(\mathcal{V})$, we denote the corresponding operator in the Liouville space representation as $\hat{\mathcal{F}}$. Using the operator transformation rule $A\rho B \iff A \otimes B^T |\rho\rangle\rangle$, we can write the above Lindblad master equation in the Liouville space $|\dot{\rho}\rangle\rangle = \hat{\mathcal{L}}|\rho\rangle\rangle$, where

$$\hat{\mathcal{L}} = \hat{\mathcal{H}} + \hat{\mathcal{D}} = -i(H \otimes I - I \otimes H^T) + \sum_i \gamma_i \left(A_i \otimes A_i^* - \frac{1}{2}(A_i^\dagger A_i \otimes I + I \otimes A_i^T A_i^*) \right). \tag{S3}$$

We can now diagonalize the superoperator

$$\hat{\mathcal{L}}|r_\alpha\rangle\rangle = \lambda_\alpha|r_\alpha\rangle\rangle \,, \qquad\qquad \hat{\mathcal{L}}^\dagger|l_\alpha\rangle\rangle = \lambda_\alpha^*|l_\alpha\rangle\rangle \,, \tag{S4}$$

where $|r_\alpha\rangle\rangle, |l_\alpha\rangle\rangle$ are the corresponding right and left eigenvectors and λ_α is the eigenvalue. Bi-orthonormality can be imposed such that $\langle\langle l_\alpha|r_\beta\rangle\rangle = \delta_{\alpha\beta}$, where the inner product in the Liouville space is equivalent to the one defined for operators $\langle\langle A|B\rangle\rangle = \mathrm{tr}(A^\dagger B)$. The evolution of the density matrix is then given as

$$|\rho(t)\rangle\rangle = \sum_\alpha e^{\lambda_\alpha t}\langle\langle l_\alpha|\rho(0)\rangle\rangle|r_\alpha\rangle\rangle \,, \tag{S5}$$

The expectation value of an observable can be calculated as

$$\langle O \rangle = \langle\langle O|\rho(t)\rangle\rangle = \sum_\alpha e^{\lambda_\alpha t}\langle\langle l_\alpha|\rho(0)\rangle\rangle\langle\langle O|r_\alpha\rangle\rangle \,. \tag{S6}$$

We can also write the Lindblad-like form for the corresponding Liouville representation of $\hat{\mathcal{L}}^\dagger$ given by

$$\mathcal{L}^\dagger\rho = +i[H, \rho] + \sum_i \gamma_i \mathcal{D}^\dagger[A_i]\rho \,, \tag{S7}$$

where $D^\dagger[A]\rho = A^\dagger \rho A - \frac{1}{2}\{A^\dagger A, \rho\}$.

---

* nishchay.suri@nasa.gov

II. UNIFORMLY DECAYING SUBSPACES

A. General Derivation

We present the general conditions to achieve Uniformly Decaying Subspaces (UDS). As above, we let $\mathcal{W}$ be the entire (pure) state space and let $\mathcal{V}$ be a chosen subspace. The expectation value of an observable O is given as

$$\langle O \rangle = \langle\!\langle O | e^{\hat{\mathcal{L}}t} | \rho(0) \rangle\!\rangle \,, \tag{S8}$$

where $\hat{\mathcal{L}}_0 = \hat{\mathcal{H}} + \hat{\mathcal{D}}$ is a general vectorized Lindblad operator. Let $\mathcal{S}$ be a subspace of the state space $\mathcal{V}$, with space of operators $\mathscr{L}(\mathcal{S}) = \mathrm{End}(\mathcal{S})$ and projector P onto $\mathcal{S}$. We write $\mathcal{P}$ to denote the corresponding projector onto $\mathscr{L}(\mathcal{S})$, with

$$\mathcal{P}(\rho) = P\rho P \tag{S9}$$

$$\hat{\mathcal{P}}|\rho\rangle\!\rangle = P \otimes P |\rho\rangle\!\rangle \tag{S10}$$

in the operator and Liouville space representations respectively. We will use the notation $\mathscr{L}(\mathcal{S})^\perp$ to denote the orthogonal complement of $\mathscr{L}(\mathcal{S})$ in $\mathscr{L}(\mathcal{V})$ (not the full orthogonal complement in $\mathscr{L}(\mathcal{W})$). We obtain a uniformly decaying subspace if

$$\hat{\mathcal{P}}\hat{\mathcal{D}} = -\gamma\hat{\mathcal{P}} \tag{S11}$$

as operators on $\mathscr{L}(\mathcal{V})$. In other words, $\hat{\mathcal{P}}\hat{\mathcal{D}}$ is a diagonalizable operator on $\mathscr{L}(\mathcal{V})$, with eigenvalue $-\gamma$ on $\mathscr{L}(\mathcal{S})$ and eigenvalue 0 on $\mathscr{L}(\mathcal{S})^\perp$. We will often find it more intuitive to decompose this into two equalities:

$$\hat{\mathcal{P}}\hat{\mathcal{D}}\hat{\mathcal{P}} = -\gamma\hat{\mathcal{P}} \,, \tag{S12}$$

$$\hat{\mathcal{P}}\hat{\mathcal{D}} = \hat{\mathcal{P}}\hat{\mathcal{D}}\hat{\mathcal{P}} \,, \tag{S13}$$

as operators on $\mathscr{L}(\mathcal{V})$. The first tells us that $\hat{\mathcal{P}}\hat{\mathcal{D}}\hat{\mathcal{P}}$ has eigenvalue $-\gamma$ on $\mathscr{L}(\mathcal{S})$. The second tells us that $\hat{\mathcal{D}}$ maps $\mathscr{L}(\mathcal{S})^\perp$ to $\mathscr{L}(\mathcal{S})^\perp$. Viewing $\hat{\mathcal{D}}$ as a matrix, we have

$$\hat{\mathcal{D}} = \begin{bmatrix} -\gamma I & 0 \\ * & * \end{bmatrix}. \tag{S14}$$

In more mathematical language, we would say that $\hat{\mathcal{D}}$ leaves the spaces $\mathscr{L}(\mathcal{V})$ and $\mathscr{L}(\mathcal{S})^\perp$ invariant, and therefore that $\hat{\mathcal{P}}\hat{\mathcal{D}}\hat{\mathcal{P}}$ is an operator on the quotient space $\mathscr{L}(\mathcal{V})/\mathscr{L}(\mathcal{S})^\perp$ (which is isomorphic to $\mathscr{L}(\mathcal{S})$ as a vector space). Eq. (S12) is the assumption that this operator on the quotient space is a scalar.

We construct a Hamiltonian that commutes with $\mathcal{P}$, implying

$$\hat{\mathcal{H}}\hat{\mathcal{P}} = \hat{\mathcal{P}}\hat{\mathcal{H}} = \hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}} \,. \tag{S15}$$

We encode the initial state and observable in the subspace $\mathcal{S}$, taking

$$|\rho(0)\rangle\!\rangle = \hat{\mathcal{P}}|\rho(0)\rangle\!\rangle \,, \qquad\qquad |O\rangle\!\rangle = \hat{\mathcal{P}}|O\rangle\!\rangle \,. \tag{S16}$$

Using Eq. S16, we rewrite the expectation value

$$\langle O \rangle_{\mathrm{UDS}} = \langle\!\langle O | \hat{\mathcal{P}} e^{\hat{\mathcal{L}}_0 t} \hat{\mathcal{P}} | \rho(0) \rangle\!\rangle = \sum_n \frac{t^n}{n!} \langle\!\langle O | \hat{\mathcal{P}} (\hat{\mathcal{H}} + \hat{\mathcal{D}})^n \hat{\mathcal{P}} | \rho(0) \rangle\!\rangle = \langle\!\langle O | e^{\hat{\mathcal{P}}\hat{\mathcal{L}}_0 \hat{\mathcal{P}} t} | \rho(0) \rangle\!\rangle = e^{-\gamma t} \langle O \rangle_C \,, \tag{S17}$$

where $\langle O \rangle_C \equiv \langle\!\langle O | e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}} t} | \rho(0) \rangle\!\rangle$ is the expectation value of the observable in the coherent case without any dissipation. We used Eq. S13 and Eq. S15 to obtain the second last equality. We can see that the decay part factors out and the effective evolution inside the uniformly decaying subspace is coherent up to the overall scaling factor.

More generally, note that $\hat{\mathcal{P}}\hat{\mathcal{L}}_0 = \hat{\mathcal{P}}\hat{\mathcal{L}}_0\hat{\mathcal{P}}$, and thus we may allow $|\rho(0)\rangle\!\rangle \in \hat{\mathcal{V}}$. Further, rather than measuring O with support only on the computational subspace, we may measure any observable O' on $\mathcal{V}$ that commutes with P. By post-selecting for measurement values that indicate a projection into $\mathscr{L}(\mathcal{S})$, we effectively measure O.

B. Sufficient conditions using Lindblad equation

We now consider the Lindblad equation

$$\dot{\rho} = \mathcal{L}\rho = \mathcal{H}\rho + \sum_{\beta \in \mathcal{B}} \gamma_\beta \mathcal{D}[A_\beta]\rho, \tag{S18}$$

where $\mathcal{B}$ is some index set and each A_β is an operator on the state space $\mathcal{W}$. We will give sufficient conditions on the operators A_β to obtain a UDS as above. Let $\mathcal{A}$ be the associative algebra generated by the A_β and $A_\beta^\dagger A_\beta$ (note we do not generally have $A_\beta^\dagger \in \mathcal{A}$). Locate a pair of nested subspaces $\mathcal{S}^\perp \subseteq \mathcal{V}$ of $\mathcal{W}$ that are invariant subspaces for the algebra $\mathcal{A}$, and let $\mathcal{S}$ (the computational subspace) be the orthogonal complement to $\mathcal{S}^\perp$ in $\mathcal{V}$, with projector $\mathcal{P}$. We in fact add the stronger assumption that A_β maps $\mathcal{V}$ into $\mathcal{S}^\perp$. Graphically, we summarize these conditions as

$$\mathcal{V} \xrightarrow{A_\beta^\dagger A_\beta} \mathcal{V}, \qquad\qquad \mathcal{S}^\perp \xrightarrow{A_\beta^\dagger A_\beta} \mathcal{S}^\perp, \qquad\qquad \mathcal{V} \xrightarrow{A_\beta} \mathcal{S}^\perp \xrightarrow{A_\beta} \mathcal{S}^\perp. \tag{S19}$$

Equivalently, for all $|\psi\rangle \in \mathcal{V}$ and all indices β, we have:

$$PA_\beta |\psi\rangle = 0, \tag{S20}$$

$$PA_\beta^\dagger A_\beta |\psi\rangle = PA_\beta^\dagger A_\beta P |\psi\rangle. \tag{S21}$$

We also assume that for some scalar $c > 0$,

$$P\left(\sum_\beta A_\beta^\dagger A_\beta\right)P = cP. \tag{S22}$$

Given the above assumptions, we let $\overline{\gamma}$ be the average of the coefficients γ_β, and define

$$\mathcal{D} = \overline{\gamma} \sum_{\beta \in \mathcal{B}} D[A_\beta] \tag{S23}$$

(with corresponding Liouville space representation $\hat{\mathcal{D}} = \overline{\gamma} \sum_{\beta \in \mathcal{B}} \hat{D}[A_\beta]$). We will consider $\hat{\mathcal{L}}_0 = \hat{\mathcal{H}} + \hat{\mathcal{D}}$, where

$$\mathcal{H}\rho = -i[H, \rho], \tag{S24}$$

corresponding to a Hamiltonian H on $\mathcal{V}$ that commutes with $\mathcal{P}$. Writing

$$\Delta\gamma_\beta = \gamma_\beta - \overline{\gamma}, \tag{S25}$$

we have $\hat{\mathcal{L}} = \hat{\mathcal{L}}_0 + \hat{\mathcal{L}}_1$, where $\hat{\mathcal{L}}_1 = \sum_\beta \Delta\gamma_\beta \hat{D}[A_\beta]$.

It is straightforward to see that the operator $\hat{\mathcal{D}}$ satisfies the condition (S11) discussed above; we give the calculation explicitly here. Note that by (S20) and (S21),

$$\mathcal{P}D[A_\beta](\rho) = PA_\beta \rho A_\beta^\dagger P - \frac{1}{2}P\left(A_\beta^\dagger A_\beta \rho + \rho A_\beta^\dagger A_\beta\right)P \tag{S26}$$

$$= -\frac{1}{2}P\left(A_\beta^\dagger A_\beta P\rho P + P\rho P A_\beta^\dagger A_\beta\right)P. \tag{S27}$$

$$= \mathcal{P}D[A_\beta]\mathcal{P}(\rho). \tag{S28}$$

Summing the penultimate line over all β and applying (S22), we obtain

$$\mathcal{P}\mathcal{D}(\rho) = -\frac{1}{2}\overline{\gamma}\left(P\left(\sum_\beta A_\beta^\dagger A_\beta\right)P\rho P + P\rho P\left(\sum_\beta A_\beta^\dagger A_\beta\right)P\right) \tag{S29}$$

$$= -\frac{1}{2}c\overline{\gamma}(\mathcal{P}\rho\mathcal{P} + \mathcal{P}\rho\mathcal{P}) \tag{S30}$$

$$= -c\overline{\gamma}\mathcal{P}(\rho). \tag{S31}$$

This is (S11) with $\gamma = c\overline{\gamma}$. Then by the calculations of Section II A, we conclude that (S17) holds for $\hat{\mathcal{L}}_0 = \hat{\mathcal{H}} + \hat{\mathcal{D}}$ as above.

After considering some examples, we will discuss the effect of the perturbation term $\hat{\mathcal{L}}_1$. We note that the fact that $\hat{\mathcal{P}}\hat{D}[A_\beta] = \hat{\mathcal{P}}\hat{D}[A_\beta]\hat{\mathcal{P}}$ implies that

$$\hat{\mathcal{P}}\hat{\mathcal{L}}_1 = \hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}. \tag{S32}$$

C. Main example: amplitude damping channel

We now consider the example of the bosonic amplitude damping channel [S1], generalizing the qubit case considered in the main text. Let $\mathcal{W}$ be the state space of n qudits, each of dimension d. Let $a = \sum_{m \geq 0} \sqrt{m} \, |m-1\rangle\langle m|$ be the bosonic annihilation operator, with a_i denoting the annihilation operator on qudit i. We consider the Lindblad equation (S18) with $\mathcal{B} = \{1, \ldots, n\}$ and $A_i = a_i$:

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1, \ \mathcal{L}_1 = \sum_i \Delta\gamma_i D[A_i], \tag{S33}$$

where all notation is as above. (Note that $d = 2$ gives the qubit case of the main text, with $a_i = \sigma_i^-$.)

Let the *degree* of a Fock basis vector $|\alpha\rangle = |\alpha_1 \cdots \alpha_n\rangle$ be $\deg|\alpha\rangle = \sum_i \alpha_i$, and define $\mathcal{S}_k$ to be the space spanned by all Fock basis vectors of degree k. We fix a degree k and consider

$$\mathcal{V} = \bigoplus_{i=0}^{k} \mathcal{S}_i, \ \mathcal{S} = \mathcal{S}_k, \ \mathcal{S}^\perp = \bigoplus_{i=0}^{k-1} \mathcal{S}_i. \tag{S34}$$

It is easy to see that these operators satisfy the assumptions of the previous section. In particular, (S19) is immediate from the fact that a_i is degree-lowering and $a_i^\dagger a_i$ is degree-preserving. We obtain (S22) by noting that

$$a^\dagger a = \sum_{m \geq 0} m \, |m\rangle\langle m| \tag{S35}$$

and therefore, on the degree-k subspace $\mathcal{S}_k$, we have

$$\sum_i a_i^\dagger a_i \, |\alpha_1 \cdots \alpha_n\rangle = \left(\sum_i \alpha_i \right) |\alpha_1 \cdots \alpha_n\rangle = k \, |\alpha_1 \cdots \alpha_n\rangle . \tag{S36}$$

Thus we obtain (S17) with $\gamma = \overline{\gamma} \deg|\alpha\rangle$.

We further note that, since the dual-rail subspace is a subspace of $\mathcal{S}_k$ in the case $d = 2$, the above also proves the UDS conditions are satisfied for that subspace.

1. Variant: unitary transformations

We also consider the following more general variant. Recall that any Fock basis vector may be expressed (up to scalar multiple) in the form $a_{i_1}^\dagger \cdots a_{i_s}^\dagger \, |\vec{0}\rangle$, where $|\vec{0}\rangle = |0 \cdots 0\rangle$ is the vacuum state. For any $n \times n$ unitary matrix $U = (u_{ij})$, we may define a unitary transformation of the Fock space by $U a_j^\dagger U^\dagger = \sum_i u_{ij} a_j^\dagger$ and $U|\vec{0}\rangle = |\vec{0}\rangle$, so that

$$U a_{i_1}^\dagger \cdots a_{i_s}^\dagger \, |\vec{0}\rangle = (U a_{i_1}^\dagger U^\dagger) \cdots (U a_{i_s}^\dagger U^\dagger) \, |\vec{0}\rangle . \tag{S37}$$

(In photonics, these are the linear optical unitaries.) Note that we have

$$U a_j U^\dagger = \sum_i u_{ij}^* a_j, \tag{S38}$$

a sum of annihilation operators on different qubits.

We now consider the case of (S18) with $\mathcal{B} = \{1, \ldots, n\}$ and $A_i = U a_i U^\dagger$. We use the same subspaces $\mathcal{V}$, $\mathcal{S}_k$, etc., as in the previous example. Since A_i is a sum of annihilation operators, it lowers the degree of Fock basis vectors by 1, and (S22) follows as before. Note also that U preserves the degree and therefore leaves the computational subspace $\mathcal{S}_k$ of degree-k vectors invariant. Then for (S22), we have

$$\sum_i A_i^\dagger A_i \, |\alpha_1 \cdots \alpha_n\rangle = \sum_i U a_i^\dagger U^\dagger U a_i U^\dagger \, |\alpha_1 \cdots \alpha_n\rangle = \sum_i U a_i^\dagger a_i U^\dagger \, |\alpha_1 \cdots \alpha_n\rangle = \deg\alpha \, |\alpha_1 \cdots \alpha_n\rangle = k \, |\alpha_1 \cdots \alpha_n\rangle . \tag{S39}$$

We note that this example behaves very similarly to the above, but now the operators A_i are no longer required to correspond to single-qudit noise.

2. Correlated Decays

Another potential example corresponds to correlated decays, as in the Lindblad operator

$$\mathcal{L}\rho = \mathcal{H}\rho + \sum_{i \neq j} \gamma_{ij} D[a_i a_j]\rho.$$

Note that

$$\sum_{i \neq j} a_i^\dagger a_j^\dagger a_i a_j |\alpha_1 \cdots \alpha_n\rangle = \sum_{i \neq j} \alpha_i \alpha_j |\alpha_1 \cdots \alpha_n\rangle = \left(\left(\sum_i \alpha_i \right)^2 - \sum_i \alpha_i^2 \right) |\alpha_1 \cdots \alpha_n\rangle, \tag{S40}$$

so in order to satisfy (S22), we must change our computational subspace to the span of Fock basis vectors with a fixed nonzero "variance." (This distinction is irrelevant in the qubit case, since all α_i are 0 or 1.) This change gives (S22), and (S19) is easily checked.

Alternatively, one may consider

$$\mathcal{L}\rho = \mathcal{H}\rho + \sum_{i,j} \gamma_{ij} D[a_i a_j]\rho,$$

where the restriction $i \neq j$ is removed. For $|\alpha_1 \cdots \alpha_n\rangle$ in the degree-k subspace, so that $\sum_i \alpha_i = k$, this satisfies

$$\sum_{i,j} a_i^\dagger a_j^\dagger a_i a_j |\alpha_1 \cdots \alpha_n\rangle = \left(\left(\sum_i \alpha_i \right)^2 - \sum_i \alpha_i \right) |\alpha_1 \cdots \alpha_n\rangle = (k^2 - k) |\alpha_1 \cdots \alpha_n\rangle. \tag{S41}$$

Then this setting may use the usual degree-k computational subspace, and the UDS assumptions (S19), (S22) easily follow. Since this example uses the same computational subspace as the standard amplitude damping case, one can actually show that they are compatible: with the Lindblad equation

$$\mathcal{L}\rho = \mathcal{H}\rho + \sum_i \gamma_i D[a_i]\rho + \sum_{i,j} \gamma_{ij} D[a_i a_j]\rho,$$

one may define

$$\mathcal{D}' = \left(\frac{1}{n} \sum_b \gamma_b \right) \sum_i D[a_i]\rho + \left(\frac{1}{n^2} \sum_{b,d} \gamma_{b,d} \right) \sum_{i,j} D[a_i a_j]\rho$$

and find that (S17) holds for $\hat{\mathcal{L}}_0 = \hat{\mathcal{H}} + \hat{\mathcal{D}}'$.

Similar considerations allow for transformations by a unitary U as above and correlated decays between more than 2 qudits. We will observe in Section II D 3, however, that the examples based on correlated decays seem not to conveniently allow for cancellation of the first-order error.

D. Perturbative Cancellation

1. Additional assumption and its consequences

We now return to the notation and generality of Section II B, in particular (S18), (S19), (S22). In addition to these assumptions, we let T be a unitary operator that commutes with $\mathcal{P}$, with corresponding superoperator $\mathcal{T}(\rho) = T\rho T^\dagger$. We assume there exists a permutation π of the index set $\mathcal{B}$ such that

$$T A_\beta T^\dagger = A_{\pi(\beta)} \text{ for all } \beta, \tag{S42}$$

and that π acts transitively on $\mathcal{B}$. As a consequence, note that $\hat{\mathcal{T}}$ acts on the operators $D[A]$ similarly, with

$$\hat{\mathcal{T}} \hat{D}[A_\beta] \hat{\mathcal{T}}^{-1} = \hat{D}[A_{\pi(\beta)}]. \tag{S43}$$

For $\hat{\mathcal{D}}$ defined in (S23), this implies

$$\hat{\mathcal{T}} \hat{\mathcal{D}} \hat{\mathcal{T}}^{-1} = \hat{\mathcal{D}}. \tag{S44}$$

For the amplitude damping case considered in Section II C, we will take T to be the cyclic shift operator given by

$$T|\alpha_1 \cdots \alpha_n\rangle = |\alpha_n \alpha_1 \alpha_2 \cdots \alpha_{n-1}\rangle, \tag{S45}$$

so that π is the permutation $(1\,2\cdots n)$ and we have $Ta_iT^\dagger = a_{i+1}$ (indices taken modulo n).

We now consider the perturbation by $\hat{\mathcal{L}}_1$ and show that the uniformly decaying subspace is stable to first order under varying decay rates. We note that the perturbation $\mathcal{L}_1$ induces a bias in the expectation value. We will account for this by averaging over multiple different experiments, shifting by a power of T each time. Let j satisfy $0 \le j < m$, where $m = |\mathcal{B}|$ is the order of T. We implement a T^j-shifted version of the Hamiltonian, $\mathcal{T}^j(H) = T^j H T^{-j}$, which corresponds to replacing $\hat{\mathcal{H}}$ by $\hat{\mathcal{T}}^j \hat{\mathcal{H}} \hat{\mathcal{T}}^{-j}$ and $\hat{\mathcal{L}}_0$ by $\hat{\mathcal{T}}^j \hat{\mathcal{H}} \hat{\mathcal{T}}^{-j} + \hat{\mathcal{D}}$. We similarly implement shifted versions of the state and observable, $\mathcal{T}^j(\rho)$ and $\mathcal{T}^j(O)$ respectively. We observe that, by (S44),

$$\hat{\mathcal{T}}^{-j}(\hat{\mathcal{T}}^j \hat{\mathcal{H}} \hat{\mathcal{T}}^{-j} + \hat{\mathcal{D}} + \hat{\mathcal{L}}_1)\hat{\mathcal{T}}^j = \hat{\mathcal{H}} + \hat{\mathcal{D}} + \hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j = \hat{\mathcal{L}}_0 + \hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j. \tag{S46}$$

The expected value of the T^j-shifted experiment is then

$$\langle\!\langle O|\hat{\mathcal{T}}^{-j}e^{t(\hat{\mathcal{T}}^j \hat{\mathcal{H}} \hat{\mathcal{T}}^{-j} + \hat{\mathcal{D}} + \hat{\mathcal{L}}_1)}\hat{\mathcal{T}}^j|\rho(0)\rangle\!\rangle = \langle\!\langle O|e^{t(\hat{\mathcal{L}}_0 + \hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j)}|\rho(0)\rangle\!\rangle, \tag{S47}$$

the same as the unshifted case with $\hat{\mathcal{L}}_1$ replaced by $\hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j$. We note that by (S43) and the definition of $\hat{\mathcal{L}}_1$,

$$\hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j = \sum_\beta \Delta\gamma_\beta \hat{\mathcal{T}}^{-j}D[A_\beta]\hat{\mathcal{T}}^j = \sum_\beta \Delta\gamma_\beta D[A_{\pi^{-j}(\beta)}] = \sum_\beta \Delta\gamma_{\pi^j(\beta)}D[A_\beta]. \tag{S48}$$

In particular, recalling that π transitively permutes the indices $\beta \in \mathcal{B}$ and that

$$\sum_\beta \Delta\gamma_\beta = \sum_\beta (\gamma_\beta - \overline{\gamma}) = \sum_\beta \gamma_\beta - m\overline{\gamma} = 0,$$

we observe that

$$\sum_{j=0}^{m-1} \hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j = \sum_\beta \left(\sum_{j=0}^{m-1} \Delta\gamma_{\pi^j(\beta)}\right)D[A_\beta] = \sum_\beta \left(\sum_{\beta' \in \mathcal{B}} \Delta\gamma_{\beta'}\right)D[A_\beta] = 0. \tag{S49}$$

2. First order error cancellation

Let us briefly return to the unshifted case. We work in the Liouville space

$$|\dot{\rho}(t)\rangle\!\rangle = (\hat{\mathcal{L}}_0 + \hat{\mathcal{L}}_1)|\rho(t)\rangle\!\rangle. \tag{S50}$$

We consider the interaction picture where we define $|\rho_I(t)\rangle\!\rangle \equiv e^{-\hat{\mathcal{L}}_0 t}|\rho(t)\rangle\!\rangle$ and $\hat{\mathcal{L}}_1^I(t) \equiv e^{-\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1 e^{\hat{\mathcal{L}}_0 t}$ and subsequently obtain the equation of motion

$$|\dot{\rho}_I(t)\rangle\!\rangle = \hat{\mathcal{L}}_1^I(t)|\rho_I(t)\rangle\!\rangle. \tag{S51}$$

We can now perturbatively solve the above equation as

$$|\rho_I(t)\rangle\!\rangle = |\rho(0)\rangle\!\rangle + \int_0^t dt'\hat{\mathcal{L}}_1^I(t')|\rho(0)\rangle\!\rangle + \int_0^t dt' \int_0^{t'} dt''\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle. \tag{S52}$$

where we used $|\rho_I(0)\rangle\!\rangle = |\rho(0)\rangle\!\rangle$. Applying $e^{\hat{\mathcal{L}}_0 t}$, we obtain

$$|\rho(t)\rangle\!\rangle = e^{\hat{\mathcal{L}}_0 t}|\rho(0)\rangle\!\rangle + \int_0^t dt'e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')|\rho(0)\rangle\!\rangle + \int_0^t dt' \int_0^{t'} dt''e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle. \tag{S53}$$

Therefore the expectation value of the observable O now is

$$\langle O\rangle_{\text{UDS}} = \langle\!\langle O|\rho(t)\rangle\!\rangle = \langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\left(|\rho(0)\rangle\!\rangle + \int_0^t dt'\hat{\mathcal{L}}_1^I(t')|\rho(0)\rangle\!\rangle + \int_0^t dt' \int_0^{t'} dt''e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle\right) \tag{S54}$$

$$= e^{-c\overline{\gamma}t}\langle O\rangle_C + \langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\left(\int_0^t dt'\hat{\mathcal{L}}_1^I(t')|\rho(0)\rangle\!\rangle + \int_0^t dt' \int_0^{t'} dt''e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle\right), \tag{S55}$$

where we apply (S17) in the last line. We will consider the 0th and 1st order terms, observing that averaging over all $\hat{\mathcal{T}}$-shifts leaves the 0th order invariant while cancelling the 1st order terms. We will consider the problem of bounding the 2nd order terms in Section II D 4.

Since $\hat{\mathcal{L}}_0$ is unaffected by $\hat{\mathcal{T}}$-shifting (up to the change in encoding of the Hamiltonian), the constant term $e^{-c\bar{\gamma}t}\langle O \rangle_C$ is unaffected. Also recall $\hat{\mathcal{L}}_1^I(t) \equiv e^{-\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1 e^{\hat{\mathcal{L}}_0 t}$, so the $\hat{\mathcal{T}}^j$-shifted version is $e^{-\hat{\mathcal{L}}_0 t}\hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j e^{\hat{\mathcal{L}}_0 t}$. Then the expected value averaged over all $\hat{\mathcal{T}}$-shifts, denoted by $\langle\!\langle O \rangle\!\rangle_{\mathrm{UDS}}$, satisfies

$$\langle\!\langle O \rangle\!\rangle_{\mathrm{UDS}} - e^{-c\bar{\gamma}t}\langle O \rangle_C = \frac{1}{m}\sum_{j=0}^{m-1}\langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\int_0^t dt' e^{-\hat{\mathcal{L}}_0 t}\hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j e^{\hat{\mathcal{L}}_0 t}|\rho(0)\rangle\!\rangle + \text{2nd order} \tag{S56}$$

$$= \frac{1}{m}\langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\int_0^t dt' e^{-\hat{\mathcal{L}}_0 t}\left(\sum_{j=0}^{m-1}\hat{\mathcal{T}}^{-j}\hat{\mathcal{L}}_1\hat{\mathcal{T}}^j\right)e^{\hat{\mathcal{L}}_0 t}|\rho(0)\rangle\!\rangle + \text{2nd order} \tag{S57}$$

$$= 0 + \text{2nd order}, \tag{S58}$$

where we apply (S49) in the last line. In particular, we were able to negate the effect of first order deviations in decay rates even without the knowledge of perturbations. Therefore, we see that uniformly decaying subspaces can be made stable up to first order.

3. Amplitude damping channel examples

We now return to the examples related to the amplitude damping channel, discussed in Section II C. As stated above, in the standard case $A_i = a_i$ we take T to be the cyclic shift operator given by

$$T|\alpha_1\cdots\alpha_n\rangle = |\alpha_n\alpha_1\alpha_2\cdots\alpha_{n-1}\rangle, \tag{S59}$$

so that $Ta_iT^\dagger = a_{i+1}$ (indices taken modulo n).

More generally, we recall the variant with $A_i = Ua_iU^\dagger$ where U is an $n \times n$ unitary matrix. We note that

$$(UTU^\dagger)A_i(UTU^\dagger)^\dagger = UTa_iT^\dagger U^\dagger = Ua_{i+1}U^\dagger = A_{i+1},$$

where we continue to take the indices of the A and a operators modulo n. Then in the U-transformed setting, we must replace the translation T with $UTU^\dagger$.

For the case of correlated decays, $A_{ij} = a_ia_j$, there is in general no operator T satisfying the desired properties. In principle, one can simply replace the role of $1, T, T^2, \ldots$ with the entire group of permutations of n qudits, S_n, and similar calculations to the above will work. But of course this would require one to perform $n!$ different experiments, leading to infeasible scaling.

4. Second order error bound

We now consider the last term in (S55). We have shown that under the assumptions of the previous section, the middle (first order) term can be cancelled. Thus if one wants to estimate $\langle O \rangle_C \approx e^{c\bar{\gamma}t}\langle\!\langle O \rangle\!\rangle_{\mathrm{UDS}}$, the error is given by the last term of (S55). To obtain this bound, we follow the techniques of the supplemental material of [S2].

Since $\langle\!\langle O| = \langle\!\langle O|\hat{\mathcal{P}}$, $\hat{\mathcal{P}}\hat{\mathcal{L}}_i = \hat{\mathcal{P}}\hat{\mathcal{L}}_i\hat{\mathcal{P}}$, and $\hat{\mathcal{P}}\hat{\mathcal{L}}_0\hat{\mathcal{P}} = \hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}} - c\bar{\gamma}\hat{\mathcal{P}}$, we have

$$\langle\!\langle O|e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle = \langle\!\langle O|\hat{\mathcal{P}}e^{\hat{\mathcal{L}}_0 t}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{L}}_1^I(t'')|\rho_I(t'')\rangle\!\rangle \tag{S60}$$

$$= \langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{L}}_0\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t'')\hat{\mathcal{P}}|\rho_I(t'')\rangle\!\rangle \tag{S61}$$

$$= e^{-c\bar{\gamma}t}\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t'')\hat{\mathcal{P}}|\rho_I(t'')\rangle\!\rangle. \tag{S62}$$

We now apply the mean value theorem: there exist $\zeta', \zeta'' \in [0, t]$ such that

$$\int_0^t dt' \int_0^{t'} dt'' e^{-c\bar{\gamma}t}\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t'')\hat{\mathcal{P}}|\rho_I(t'')\rangle\!\rangle = \frac{t^2}{2}e^{-c\bar{\gamma}t}\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle. \tag{S63}$$

We briefly recall the Schatten p-norms, for $1 \le p \le \infty$, $\|T\|_p = \left(\text{Tr}\left(\sqrt{T^\dagger T}^p\right)\right)^{1/p}$. For $p = \infty$, this is the standard operator norm. Rewriting the above as a 1-norm and applying Hölder's inequality for Schatten p-norms, we have

$$\frac{t^2}{2}e^{-c\bar{\gamma}t}\left|\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle\right| = \frac{t^2}{2}e^{-c\bar{\gamma}t}\|\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle\|_1 \tag{S64}$$

$$\le \frac{t^2}{2}e^{-c\bar{\gamma}t}\|\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\|_\infty \cdot \|\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle\|_1. \tag{S65}$$

Since $e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}$ is unitary, we have $\|\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\|_\infty = \|O\|_\infty = \|O\|_{op}$, the standard operator norm of O. Then we must only bound

$$\|\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle\|_1 = \|\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(\zeta'')\hat{\mathcal{P}}|\rho_I(\zeta'')\rangle\!\rangle\|_1 \tag{S66}$$

$$= \|e^{-\hat{\mathcal{P}}\hat{\mathcal{L}}_0\hat{\mathcal{P}}\zeta'}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}e^{\hat{\mathcal{P}}\hat{\mathcal{L}}_0\hat{\mathcal{P}}(\zeta'-\zeta'')}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}|\rho(t'')\rangle\!\rangle\|_1 \tag{S67}$$

$$= e^{-c\bar{\gamma}(-\zeta'+(\zeta'-\zeta''))}\|e^{-\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}\zeta'}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}(\zeta'-\zeta'')}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}|\rho(\zeta'')\rangle\!\rangle\|_1 \tag{S68}$$

$$\le e^{c\bar{\gamma}\zeta''}\|e^{-\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}\zeta'}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}(\zeta'-\zeta'')}\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}\|_{op,1} \tag{S69}$$

$$= e^{c\bar{\gamma}\zeta''}\|\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}\|_{op,1}^2, \tag{S70}$$

where $\|\cdot\|_{op,1}$ is the operator norm with respect to the Schatten 1-norm, and we use the fact that operator norms are unitarily invariant. In total, then, the error term of interest satisfies

$$\left|\int_0^t dt' \int_0^{t'} dt'' e^{-c\bar{\gamma}t}\langle\!\langle O|e^{\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}}t}\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t')\hat{\mathcal{P}}\hat{\mathcal{L}}_1^I(t'')\hat{\mathcal{P}}|\rho_I(t'')\rangle\!\rangle\right| \le \frac{t^2}{2}e^{-c\bar{\gamma}(t-\zeta'')}\|O\|_{op} \cdot \|\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}\|_{op,1}^2 \tag{S71}$$

$$\le \frac{t^2}{2}\|O\|_{op} \cdot \|\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}\|_{op,1}^2. \tag{S72}$$

The analogous result holds in the T-shifted case. Combining this with the results of Section II D, we obtain

$$\left|\langle\!\langle O\rangle\!\rangle_{\text{UDS}} - e^{-c\bar{\gamma}t}\langle O\rangle_C\right| \le \frac{t^2}{2}\|O\|_{op} \cdot \|\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}\|_{op,1}^2. \tag{S73}$$

5. Return to amplitude damping example

We return to the example of Section II C, with qudits of dimension d, $A_i = a_i$ the annihilation operator on the ith qudit, and computational subspace the degree-k subspace $\mathcal{S}_k$. The bound above is determined by the operator norm of $\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}$. We note

$$\mathcal{P}\mathcal{L}_1\mathcal{P}(\rho) = P\mathcal{L}_1(P\rho P)P = -\frac{1}{2}\sum_i \Delta\gamma_i\{a_i^\dagger a_i, \rho\} = -\frac{1}{2}\sum_i \Delta\gamma_i\left\{\sum_s s\,|s\rangle\langle s|, \rho\right\}.$$

In particular, $\hat{\mathcal{P}}\hat{\mathcal{L}}_1\hat{\mathcal{P}}$ is diagonalizable in the Fock basis $|\alpha_1\cdots\alpha_n\rangle\langle\alpha_1'\cdots\alpha_n'|$, so its operator norm is the modulus of the largest eigenvalue. One can easily verify that this largest eigenvalue is bounded above by $k\lambda$, where $\lambda = \max_i|\Delta\gamma_i|$. This gives the bound

$$\left|\langle\!\langle O\rangle\!\rangle_{\text{UDS}} - e^{-k\bar{\gamma}t}\langle O\rangle_C\right| \le \frac{(tk\lambda)^2}{2}\|O\|_{op}. \tag{S74}$$

The same result holds in the case $A_i = Ua_iU^\dagger$.

E. Circuit case

We now discuss why one can apply the techniques discussed above, specifically the cancellation of errors up to first order, to the discretized example at the end of the main text.

We let $\mathcal{E} = \mathcal{E}(\eta)$ be a single-qudit discretized amplitude damping channel, given by the following Kraus operators $K_s(\eta)$ for $s \geq 0$ and a parameter η [S1] (where $\eta = 1 - \gamma$ in ibid.):

$$K_s(\eta) = \sum_r \sqrt{\binom{r}{s}} \eta^{r-s} (1 - \eta)^s |r - s\rangle\langle r|. \tag{S75}$$

We write $\mathcal{E}_i(\eta_i)$ for $\mathcal{E}$ acting on the ith qudit with parameter $\eta = \eta_i$.

Now suppose we prepare a state $\rho \in \mathcal{L}(\mathcal{V})$, apply unitaries $U_1, \ldots, U_{N_T}$ on $\mathcal{V}$ that commute with P (with corresponding operators $\hat{\mathcal{U}}_1, \ldots, \hat{\mathcal{U}}_{N_T}$ on the Liouville space), and measure an observable O with $POP = O$. We model dissipation in a discretized manner by replacing each $\hat{\mathcal{U}}_j$ by $(\otimes_i \hat{\mathcal{E}}_i(\eta_i))\hat{\mathcal{U}}_j$. The expectation value is the following:

$$\langle O \rangle = \langle\!\langle O | \hat{\mathcal{P}} (\bigotimes_i \hat{\mathcal{E}}_i(\eta_i)) \hat{\mathcal{U}}_1 (\bigotimes_i \hat{\mathcal{E}}_i(\eta_i)) \cdots (\bigotimes_i \hat{\mathcal{E}}_i(\eta_i)) \hat{\mathcal{U}}_{N_T} | \rho \rangle\!\rangle. \tag{S76}$$

Since all K_s preserve or lower the degree, keeping $\mathcal{L}(\mathcal{S})^\perp$ and $\mathcal{L}(\mathcal{V})$ invariant, we have $\hat{\mathcal{P}}\hat{\mathcal{E}}_i = \hat{\mathcal{P}}\hat{\mathcal{E}}_i\hat{\mathcal{P}}$. Further, the $\hat{\mathcal{U}}_i$ commute with $\hat{\mathcal{P}}$ by assumption. Therefore,

$$\langle O \rangle = \langle\!\langle O | (\bigotimes_i \hat{\mathcal{P}}\hat{\mathcal{E}}_i(\eta_i)\hat{\mathcal{P}}) \hat{\mathcal{U}}_1 (\bigotimes_i \hat{\mathcal{P}}\hat{\mathcal{E}}_i(\eta_i)\hat{\mathcal{P}}) \cdots (\bigotimes_i \hat{\mathcal{P}}\hat{\mathcal{E}}_i(\eta_i)\hat{\mathcal{P}}) \hat{\mathcal{U}}_{N_T} | \rho \rangle\!\rangle. \tag{S77}$$

Since $POP = O$ (meaning that we post-select for measurements in the degree-k subspace), and $|\rho\rangle\!\rangle$ involves only terms of degree k or lower, all terms involving the degree-raising operators K_s for $s > 0$ must vanish. Then only the K_0 term of each $\hat{\mathcal{P}}\hat{\mathcal{E}}_i\hat{\mathcal{P}}$ contributes to the expectation. Letting $\hat{K}_0^{(i)}(\eta)$ be the corresponding operator on $\mathcal{L}(\mathcal{V})$, we note that we may replace $\otimes_i \hat{\mathcal{P}}\hat{\mathcal{E}}_i(\eta_i)\hat{\mathcal{P}}$ with

$$\hat{\mathscr{E}}(\eta_1, \ldots, \eta_n) \equiv \bigotimes_i \hat{\mathcal{P}}\hat{K}_0^{(i)}(\eta_i)\hat{\mathcal{P}}, \tag{S78}$$

and write

$$\langle O \rangle = \langle\!\langle O | \hat{\mathscr{E}}(\eta_1, \ldots, \eta_n) \hat{\mathcal{U}}_1 \hat{\mathscr{E}}(\eta_1, \ldots, \eta_n) \cdots \hat{\mathscr{E}}(\eta_1, \ldots, \eta_n) \hat{\mathcal{U}}_{N_T} | \rho \rangle\!\rangle. \tag{S79}$$

We will show that the first-order cancellation observed above extends to this discretized setting. Toward that end, for comparison with the continuous case, we take $\eta_i = \exp(-\Delta t \gamma_i)$, where γ_i is the parameter in the Lindblad master equation (S33) and $\Delta t = t/N_T$ is the time elapsed per unitary (here t is the total time elapsed, as above). We define

$$\bar{\eta} = (\eta_1 \cdots \eta_n)^{1/n} = \exp\left(-\frac{1}{n}\Delta t \sum_i \gamma_i\right) = \exp(-\Delta t \bar{\gamma}) \text{ and } \Delta\eta_i = \eta_i/\bar{\eta} = \exp(-\Delta t \Delta\gamma_i), \tag{S80}$$

and note that $\hat{K}_0^{(i)}(\eta_i) = \hat{K}_0^{(i)}(\bar{\eta})\hat{K}_0^{(i)}(\Delta\eta_i)$, so that

$$\hat{\mathscr{E}}(\eta_1, \ldots, \eta_n) = \hat{\mathscr{E}}(\bar{\eta}, \ldots, \bar{\eta})\hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n). \tag{S81}$$

For $|\rho'\rangle\!\rangle$ in the degree-k subspace, we note that

$$\hat{\mathscr{E}}(\bar{\eta}, \ldots, \bar{\eta})|\rho'\rangle\!\rangle = \bar{\eta}^k = \exp(-k\Delta t \bar{\gamma}), \tag{S82}$$

so we obtain

$$\langle O \rangle = e^{-k N_T \Delta t \bar{\gamma}} \langle\!\langle O | \hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n) \hat{\mathcal{U}}_1 \hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n) \cdots \hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n) \hat{\mathcal{U}}_{N_T} | \rho \rangle\!\rangle, \tag{S83}$$

where we note that $N_T \Delta t = t$ is the total time.

Now, we consider a single expression $\hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n)|\rho'\rangle\!\rangle$ for $\rho' \in \mathcal{L}(\mathcal{V})$. As above, it suffices to consider $\rho' \in \mathcal{L}(\mathcal{S})$. For $\rho' = |\alpha_1 \cdots \alpha_n\rangle\langle\alpha_1' \cdots \alpha_n'| \in \mathcal{L}(\mathcal{S})$, $\hat{\mathcal{P}}\hat{\mathcal{E}}_i(\Delta\eta_i)\hat{\mathcal{P}}$ acts by the scalar $(\Delta\eta_i)^{(\alpha_i + \alpha_i')/2}$. In particular, we see that $\hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n)$ acts by the scalar

$$\prod_i (\Delta\eta_i)^{(\alpha_i + \alpha_i')/2} = \exp\left(-\frac{1}{2}\Delta t \sum_i (\alpha_i + \alpha_i')\Delta\gamma_i\right) = 1 - \frac{1}{2}\Delta t \sum_i (\alpha_i + \alpha_i')\Delta\gamma_i + O((\Delta t \Delta\gamma_{\max})^2). \tag{S84}$$

Averaging over all cyclic permutations of the qudits as in Section II D, we see that the averaged channel acts on $\rho' = |\alpha_1 \cdots \alpha_n\rangle\langle\alpha_1' \cdots \alpha_n'| \in \mathcal{L}(\mathcal{S})$ by the scalar

$$1 - \frac{1}{2n}\Delta t \sum_i (\alpha_i + \alpha_i') \sum_j \Delta\gamma_j + O((\Delta t \Delta\gamma_{\max})^2) = 1 + O((\Delta t \Delta\gamma_{\max})^2). \tag{S85}$$

Then up to first order in Δt and $\Delta\gamma_{\max}$, the superoperator $\hat{\mathscr{E}}(\Delta\eta_1, \ldots, \Delta\eta_n)$ acts trivially. This suffices to show that

$$\langle\!\langle O \rangle\!\rangle_{\mathrm{UDS}} = e^{-kt\bar{\gamma}}\left(\langle\!\langle O|\hat{\mathcal{U}}_1 \cdots \hat{\mathcal{U}}_{N_T}|\rho\rangle\!\rangle + O((\Delta t \Delta\gamma_{\max})^2)\right) \tag{S86}$$

and

$$\langle\!\langle O \rangle\!\rangle_{\mathrm{UDS}} - e^{-c\bar{\gamma}t}\langle O \rangle_C = O((\Delta t \Delta\gamma_{\max})^2) \tag{S87}$$

as desired.

[S1] I. L. Chuang, D. W. Leung, and Y. Yamamoto, Bosonic quantum codes for amplitude damping, Physical Review A **56**, 1114 (1997).
[S2] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Physical review letters **119**, 180509 (2017).