New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
20 April 2015

REVISED
23 August 2015

ACCEPTED FOR PUBLICATION
5 October 2015

PUBLISHED
5 November 2015

PAPER

Estimating the coherence of noise

Joel Wallman[1,3], **Chris Granade**[2], **Robin Harper**[2] **and Steven T Flammia**[2]

[1] Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
[2] Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, Australia
[3] Author to whom any correspondence should be addressed.

E-mail: joel.j.wallman@gmail.com

Keywords: characterization, randomized benchmarking, coherent errors

Supplementary material for this article is available online

Abstract

Noise mechanisms in quantum systems can be broadly characterized as either coherent (i.e., unitary) or incoherent. For a given fixed average error rate, coherent noise mechanisms will generally lead to a larger worst-case error than incoherent noise. We show that the coherence of a noise source can be quantified by the *unitarity*, which we relate to the average change in purity averaged over input pure states. We then show that the unitarity can be efficiently estimated using a protocol based on randomized benchmarking that is efficient and robust to state-preparation and measurement errors. We also show that the unitarity provides a lower bound on the optimal achievable gate infidelity under a given noisy process.

To harness the advantages of quantum information processing, quantum systems have to be controlled to within some maximum threshold error. Certifying whether the error is below the threshold is possible by performing full quantum process tomography [1, 2], however, quantum process tomography is both inefficient in the number of qubits and is sensitive to state-preparation and measurement errors (SPAM) [3].

Randomized benchmarking [4–9] and direct fidelity estimation [10, 11] have been developed as efficient methods for estimating the average infidelity of noise to the identity. However, the worst-case error, as quantified by the diamond distance from the identity, can be more relevant to determining whether an experimental implementation is at the threshold for fault-tolerant quantum computation [12]. The best possible bound on the worst-case error (without further information about the noise) scales as the square root of the infidelity and can be orders of magnitude greater than the reported average infidelity [13, 14].

However, this scaling of the worst-case bound is only known to be saturated by unitary noise. If the noise is known to be stochastic Pauli noise, the worst-case error is directly proportional to the average infidelity [9], vastly improving on the general bound. Consequently, quantifying the intermediate regime between unitary and fully incoherent noise may allow the bound on the worst-case error to be substantially improved.

Randomized benchmarking is also emerging as a useful tool for diagnosing the noise in an experiment [15, 16], which can then be used to optimize the implementation of gates by varying the experimental design. In this spirit, an experimental protocol for characterizing the coherence of a noise channel will be an important tool as the quest to build a fault-tolerant quantum computer progresses.

In this paper, we present a protocol for estimating a particular quantification of the coherence of noise, which we term the *unitarity*, in the experimental implementation of a unitary 2-design. Our protocol is efficient and robust against SPAM, and is a minor modification of randomized benchmarking. The unitarity is defined as the average change in the purity of a pure state after applying the noise channel, with the contributions due to the identity component subtracted off, (see equation (4)) and is closely related to the purity of the Jamiołkowski isomorphic state (see proposition 9). We show that the unitarity is invariant under unitary gates and attains its maximal value if and only if the noise is unitary. Furthermore, we show that the unitarity can be combined with the average gate fidelity to quantify how far a noise channel is from depolarizing noise. Finally, we show that the unitarity of a noise channel provides a lower bound on the best achievable gate infidelity assuming perfect unitary control.

Our approach to quantifying coherence complements other recent work on quantifying coherence since we focus on the coherence of quantum *operations* rather than the coherence of quantum *states* relative to a preferred basis [17].

1. Defining unitarity

We begin by defining the unitarity of a noise channel $\mathcal{E}$: $\mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$, that is, a completely positive (CP) linear map that takes quantum states to quantum states. The purity of a quantum state ρ is $\mathrm{Tr}\ \rho^\dagger\rho \in [0,\ 1]$ with $\mathrm{Tr}\ \rho^\dagger\rho = 1$ if and only if ρ is a pure state. An initial candidate for a definition of the unitarity of $\mathcal{E}$ is

$$\int \mathrm{d}\psi\ \mathrm{Tr}\Big[\mathcal{E}(\psi)^\dagger \mathcal{E}(\psi)\Big], \tag{1}$$

that is, as the purity of the output states averaged over all pure state inputs. However, this definition is problematic, since it would lead to the nonunital state-preparation channel

$$\mathcal{E}_0(\rho) = \mathrm{Tr}(\rho)|0\rangle\langle 0| \tag{2}$$

having the same value of unitarity as a unitary channel, even though it does not preserve coherent superpositions. Similarly, the (trace-decreasing) filtering channel

$$\mathcal{E}_1(\rho) = |0\rangle\langle 0|\ \rho\ |0\rangle\langle 0| \tag{3}$$

does not preserve coherent superpositions and so should have the same unitarity value as a complete depolarizing channel. Both of these problematic channels arise when either the identity is mapped to coherent terms or *vice versa*.

To avoid these issues, we define the unitarity of a noise channel to be the average purity of output states, with the identity components subtracted, averaged over all pure states. That is, we define

$$u(\mathcal{E}) = \frac{d}{d-1}\int \mathrm{d}\psi\ \mathrm{Tr}\Big[\mathcal{E}'(\psi)^\dagger \mathcal{E}'(\psi)\Big], \tag{4}$$

where the normalization factor is chosen so that $u(\mathcal{I}) = 1$ and $\mathcal{E}'$ is defined so that $\mathcal{E}'(A) = \mathcal{E}(A) - [\mathrm{Tr}\ \mathcal{E}(A)/\sqrt{d}\,]\mathbb{1}$ for all traceless A (to account for trace-decreasing channels, such as in equation (2)) and $\mathcal{E}'(\mathbb{1}_d) = 0$ (to account for non-unital channels, such as in equation (3)). Equivalently, if $\{A_1, \ldots, A_{d^2}\}$ is any set of traceless and trace-orthonormal operators (e.g., the normalized Paulis), then we can define the generalized Bloch vector $\mathbf{n}(\rho)$ of a density operator ρ with unit trace to be the vector of $d^2 - 1$ expansion coefficients

$$\rho = \mathbb{1}_d/d + \sum_{k>1} n_k A_k. \tag{5}$$

Our definition of the unitarity is then equivalent to

$$u(\mathcal{E}) = \frac{d}{d-1}\int \mathrm{d}\psi\ \Big\|\mathbf{n}[\mathcal{E}(\psi)] - \mathbf{n}\Big[\mathcal{E}\Big(\mathbb{1}_d/d\Big)\Big]\Big\|^2, \tag{6}$$

that is, the average squared length (i.e., Euclidean norm) of the generalized Bloch vector after applying the map $\mathcal{E}$ with the component due to the identity subtracted off.

2. The estimation protocol

We now present a protocol for characterizing the unitarity of the noise in an experimental implementation of a unitary 2-design $\mathcal{G}$ under the assumption that the experimental implementation of any $U \in \mathcal{G}$ can be written as $\mathcal{U} \circ \mathcal{E}$ where $\mathcal{U}$ denotes the channel corresponding to conjugation by U and $\mathcal{E}$ is a completely positive, trace-preserving (CPTP) channel independent of U. (Note that, as in all randomized benchmarking papers, the assumption that $\mathcal{E}$ is independent of $\mathcal{U}$ can be relaxed without dramatically effecting the results [8, 13, 16].)

The protocol is to repeat many independent trials of the following.

- Choose a sequence $\mathbf{j} = (j_1, \ldots, j_m)$ of m integers in $\mathbb{N}_{|\mathcal{G}|} = \{1, \ldots, |\mathcal{G}|\}$ uniformly at random.
- Estimate the expectation value $Q_\mathbf{j}$ of an operator Q after preparing the state ρ and applying the sequence $U_\mathbf{j} = U_{j_m} U_{j_{m-1}} \ldots U_{j_1}$ of operators. In the ideal case that $\mathcal{E} = \mathcal{I}$, the expectation value is given by

$$Q_\mathbf{j} = \mathrm{Tr}\Big(QU_\mathbf{j}\ \rho U_\mathbf{j}^\dagger\Big). \tag{7}$$

We will show in section 4, theorem 5 that, under the above assumptions on the noise, the expected value of $Q_{\mathbf{j}}^2$ over all random sequences $\mathbf{j}$ obeys

$$\mathbb{E}_{\mathbf{j}}\left[Q_{\mathbf{j}}^2\right] = A + Bu(\mathcal{E})^{m-1} \tag{8}$$

for trace-preserving noise, where A and B are constants incorporating SPAM and the nonunitality of the noise and $u(\mathcal{E}) \in [0,\ 1]$ is the unitarity of the noise defined in equation (4), with $u(\mathcal{E}) = 1$ if and only if $\mathcal{E}$ is unitary.

Therefore estimating $\mathbb{E}_{\mathbf{j}}[Q_{\mathbf{j}}^2]$ for multiple values of m using the above protocol and fitting to equation (8) gives an efficient and robust estimator of the unitarity.

2.1. Estimators

Note that, as opposed to standard presentations of randomized benchmarking, we are considering the expectation of an operator Q rather than the probability of a single outcome. These descriptions are mathematically equivalent, though expressing the measurement as an observable allows a more concise description of measurement procedures that involve averaging over multiple observables. For example, we can average over the non-identity Pauli operators, while keeping the sequence the same. As will be discussed in section 4, this allows us to simulate a two-state measurement involving S, the SWAP operator. We term this the *purity measurement*, as it estimates the relevant state-dependent term in

$$\mathrm{Tr}\left(\rho_{\mathbf{j}}^2\right) = \frac{1}{d} + \|\mathbf{n}(\rho_{\mathbf{j}})\|^2, \tag{9}$$

the purity of the state $\rho_{\mathbf{j}}$ produced by the sequence $\mathbf{j}$. What we actually use is a shifted and rescaled version of this defined by

$$P_{\mathbf{j}} = \frac{d}{d-1}\|\mathbf{n}(\rho_{\mathbf{j}})\|^2, \tag{10}$$

which for physical states is always in the interval $[0,\ 1]$. For a single qubit, and measuring in the Pauli basis, this quantity is just $P_{\mathbf{j}} = \langle X\rangle^2 + \langle Y\rangle^2 + \langle Z\rangle^2$, where each expectation value is taken with respect to the state $\rho_{\mathbf{j}}$.

The purity measurement can be performed in one of two ways. The direct way involves using two copies of the experiment (with the same sequence) that are run in parallel and a SWAP gate applied immediately prior to measurement. A method using only one copy makes use of the expansion

$$S = \sum_k A_k \otimes A_k^\dagger \tag{11}$$

for any orthonormal operator basis $\{A_k\}$ (e.g., the normalized Paulis) by adding up the expectation values over measurements in the operator basis for the same sequences, that is, by estimating $\sum_k \mathbb{E}_{\mathbf{j}}[(A_k)_{\mathbf{j}}^2]$.

Implementing the purity measurement using this averaging reduces the between-sequence contribution to the uncertainty in our estimates of $\mathbb{E}_{\mathbf{j}}[Q_{\mathbf{j}}^2]$, since if the noise is approximately unitary, then the final state will be relatively pure but will generically overlap with all non-identity Paulis. We note, however, that the above summation over a trace-orthonormal basis is not scalable with the number of qubits, since there are exponentially many n-qubit Paulis. We leave possible optimizations and an analysis of the scalable two-copy protocol as an open problem.

Also note that unlike in standard randomized benchmarking, we do not require the unitary 2-design to be a group since we do not require an inverse operation, or even that the set $\mathcal{G}$ is closed under composition.

2.2. Trace-decreasing noise

More generally, some experimental noise $\mathcal{E}$ may be trace-decreasing with an average survival rate

$$S(\mathcal{E}) = \int \mathrm{d}\psi\ \mathrm{Tr}[\mathcal{E}(\psi)], \tag{12}$$

which is the amount of the trace of the quantum state ψ that survives the error channel $\mathcal{E}$, averaged over the Haar measure $\mathrm{d}\psi$. When $\mathcal{E}$ is itself the average noise over $\mathcal{G}$, the average loss rate can be estimated by

$$\mathbb{E}_{\mathbf{j}}\left[Q_{\mathbf{j}}\right] = CS(\mathcal{E})^{m-1} \tag{13}$$

where C is a constant determined by SPAM [16].

For trace-decreasing noise, the standard decay curve in equation (8) can be generalized to

$$\mathbb{E}_{\mathbf{j}}\left[Q_{\mathbf{j}}^2\right] = A\lambda_+^{m-1} + B\lambda_-^{m-1}, \tag{14}$$

for some constants A and B where

$$\lambda_+ + \lambda_- = S(\mathcal{E})^2 + u(\mathcal{E}). \tag{15}$$

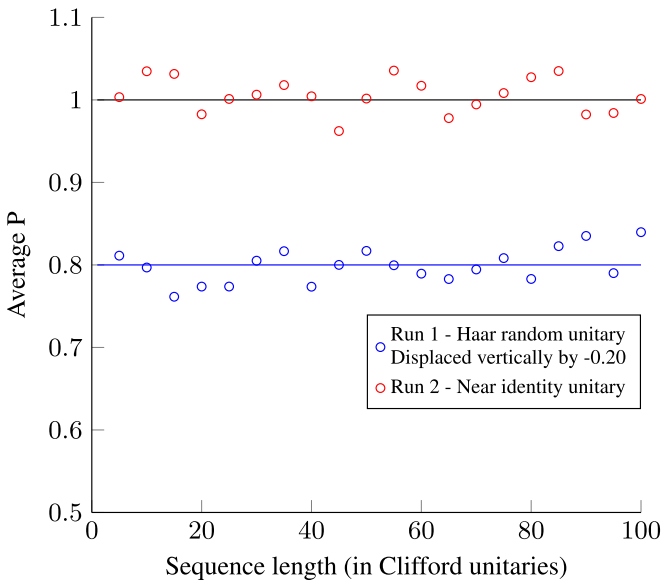

Figure 1. Comparison of our model (equation (8)) with simulated data, where the error channel is taken to be a fixed unitary drawn from the Haar measure (blue), or given by a rotation of 0.1 about σ_x (red). In the simulated data, the purity is measured following the formula of equation (10). The solid lines show the model equation (8) for the calculated values of A, B and $u(\mathcal{E})$ for each scenario. This shows that even for extreme values of unitary noise, our model correctly predicts the unitarity.

The above protocol is a variation of standard randomized benchmarking experiments, and is very similar to the protocol for estimating loss presented in [16]. In particular, one estimates an exponential decay rate in an exactly analogous manner (see equation (8)) and the result is obtained in a manner that is robust to SPAM.

However, there are three small but crucial differences to the experimental protocol presented in [16], leading to significant differences in the analysis and interpretation of the decay curves. Most importantly, the post-processing is different, since in the present paper the survival probabilities for the individual sequences are squared before they are averaged. Secondly, the preparation and measurement procedures in the loss protocol of [16] are ideally the maximally mixed state and the trivial (identity) measurement respectively, which is outperformed in this protocol by the use of the purity measurement. Finally, the current protocol requires a unitary 2-design, whereas the loss protocol only requires a unitary 1-design (although it also works for a unitary 2-design).

3. Numerical simulations

We now illustrate our experimental protocol by numerically simulating it for a variety of single-qubit noise models and showing that the results fit well to (8). In figure 1, we give an example of the correctness of our model equation (8) by showing that it agrees with simulated experimental data in the extreme case that the error channel is a fixed unitary. In figure 2, we demonstrate the utility of equation (8) as an estimation model by estimating $u(\mathcal{E})$ from simulated data drawn according to our protocol. We simulate measurement error on each measurement with small independent random orthogonal matrices, scaling the unital components with a random factor between 0.95 and 1.0. In both these simulations we simulate SPAM on the prepared state by applying a random near-identity unitary. We choose $\mathcal{G}$ to be the single-qubit Clifford group.

Concretely, in figure 1 we show two runs. In the first we set $\mathcal{E}$ to be some fixed (systematic) unitary chosen randomly according to the Haar measure (*Haar-random unitary*) and some near-identity unitary represented by a rotation of 0.1 radians around the X-axis of the Bloch sphere (*near-identity unitary*). As figure 1 demonstrates, unitary noise results in no decay in the number of sequences.

In figure 2, we show different types of unital noise composed with the nonunital amplitude-damping channel

$$\mathcal{E}_{\mathrm{d}}(\rho) = p\,|0\rangle\langle 0| + (1-p)\rho \tag{16}$$

to simulate relaxation to a ground state. The particular unital channels we consider are a Haar-random unitary and a gate-dependent noise channel corresponding to choosing a fixed perturbation of the eigenvalues of a unitary g by $e^{i\epsilon}$ to simulate over/under-rotation errors, where the perturbations ϵ are chosen independently and uniformly from $[-0.1,\ 0.1]$ radians for each gate (*rotation channel*).

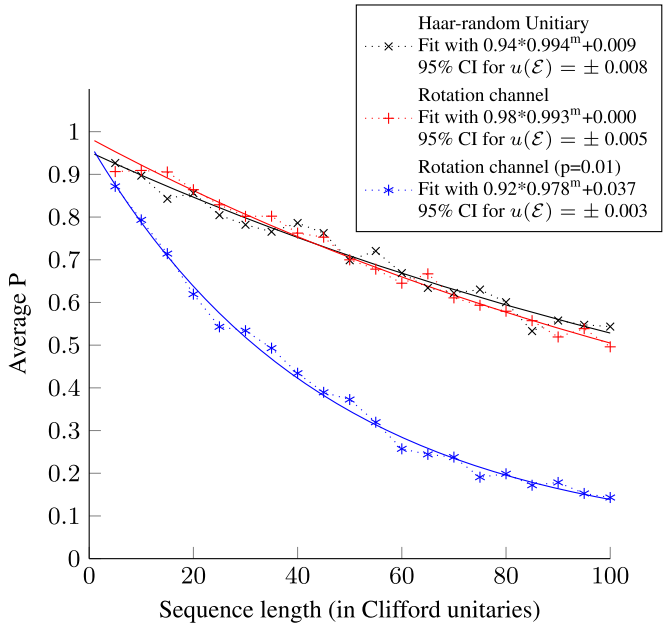

Figure 2. Numerical purity decay curve for our protocol when the noise consists of the non-unital channel in equation (16) with $p = 0.003$ composed with: a Haar-random unitary (black crosses), and fixed gate-dependent over/under-rotations where the eigenvalues for each gate are perturbed by $e^{\pm i\delta}$ for some $\delta \in [-0.01, 0.01]$ chosen independently and uniformly (red pluses). The purity is measured in the same manner as in figure 1. The third plot (blue stars) shows (gate-dependent) over/under-rotations composed with the nonunital channel with $p = 0.01$. The solid lines give the fit to equation (8), where the slope gives an estimate for $u(\mathcal{E})$ given by the values 0.994, 0.993 and 0.978, consistent with the theoretical values of 0.994, 0.994 and 0.981, respectively. Confidence intervals for $u(\mathcal{E})$ were calculated assuming Gaussian noise, such that the least-squares fit is a maximum likelihood estimator.

Note that the statistical fluctuations in figures 1 and 2 arise from between-sequence variations and within-sequence variations. The between-sequence variations arise from sampling a small number of random sequences (30 sequences in this case) relative to the total number. The between-sequence variations are minimised by measuring an observable for the purity. A perturbation expansion of the form $\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{I}} - r\boldsymbol{\delta}$ (where r is the average gate infidelity of $\boldsymbol{\mathcal{E}}$ to the identity) together with appropriate bounds on the diamond norm can be used to bound these fluctuations and show that they must decrease with gate infidelity, as in [13]. However, a more detailed analysis is complicated by the complexity of the relevant representation theory (that is, four-fold tensor products). The within-sequence variations arise from the need to estimate the expectation values of the observables. For the purpose of figures 1 and 2, we used an unbiased estimator of the squared expectation values, simulating N measurements (with N set to 150).

Finally, we consider the unitarity of random channels drawn from the random ensemble of Bruzda *et al* [18], using the QuTiP software package [19] to draw channels and compute their unitarity (see supplemental material). As shown in figure 3, the distribution of unitarities depends strongly on the Kraus rank of the random channel. Moreover, as demonstrated in figure 4, this information is correlated with, but distinct from, the average gate fidelity.

4. Derivation of the fit models

We now derive the decay curve in equation (14) for trace-non-increasing noise and show how the decay curve in equation (8) emerges as a special case for trace-preserving noise.

Since we are dealing with sequences of channels, it will be convenient to work in the Liouville representation. Since a quantum channel is a linear map between finite-dimensional vector spaces, it is always possible to represent it as a matrix acting on basis coefficients in some given bases for the vector spaces.

In order to construct the Liouville representation of channels, let $\mathcal{A} = \{A_1, \ldots, A_{d^2}\}$ be an orthonormal basis of $\mathbb{C}^{d\times d}$ according to the Hilbert–Schmidt inner product $\langle A, B\rangle = \mathrm{Tr}\, A^\dagger B$. Any density matrix ρ can be expanded as $\rho = \sum_{k\in\mathbb{N}_{d^2}} \langle A_k, \rho\rangle A_k$ and so we can identify ρ with a column vector $|\rho) \in \mathbb{C}^{d^2}$ whose kth entry is $\langle A_k, \rho\rangle$. The Liouville representation of a channel $\mathcal{E}$ is then the unique matrix $\boldsymbol{\mathcal{E}} \in \mathbb{C}^{d^2\times d^2}$ such that $\boldsymbol{\mathcal{E}}\,|\rho) = |\mathcal{E}[\rho])$, which has entries $\boldsymbol{\mathcal{E}}_{kl} = \langle A_k, \mathcal{E}(A_l)\rangle = (A_k\,|\boldsymbol{\mathcal{E}}|\,A_l)$. An immediate consequence of the

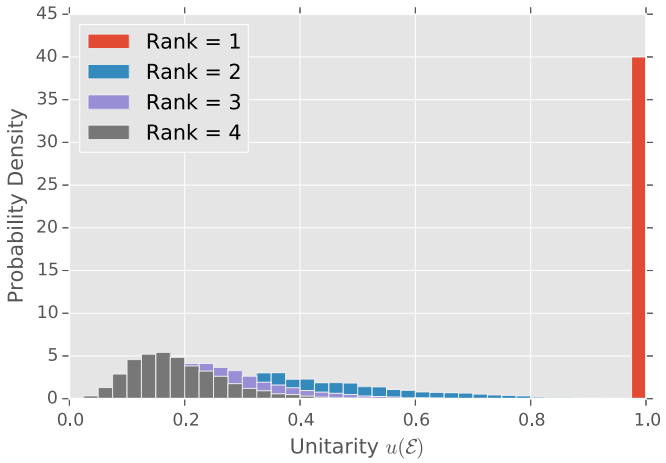

Figure 3. Unitarity of single-qubit CPTP channels chosen according to the random distributions of Bruzda *et al* [18] with varying ranks of the Kraus operators, demonstrating that the unitarity carries information about the structure of the channel. In particular, channels which require more Kraus operators to specify tend towards much smaller unitarity.

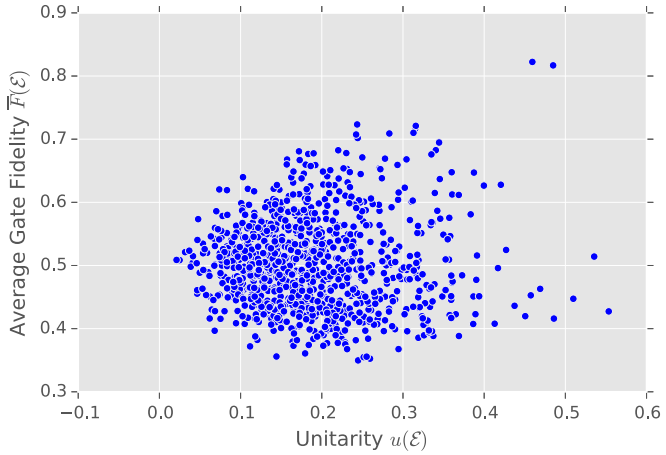

Figure 4. Unitarity of single-qubit CPTP channels chosen according to the random distributions of Bruzda *et al* [18], plotted versus their fidelities to the identity channel. This example shows that even though the two quantities are correlated, they are not redundant and give different insight into the structure of the noise.

uniqueness of $\boldsymbol{\mathcal{E}}$ is that the composition of abstract maps is represented in the Liouville representation by matrix multiplication.

The Liouville representation of unitary channels forms a unitary projective representation of the unitary group $U(d)$. When we wish to emphasize the Liouville representation as a formal representation (rep) of the unitary group $U(d)$ (or subgroups thereof), we will use the notation $\phi_L(U)$ instead of $\boldsymbol{\mathcal{U}}$. With this notation, it is easy to verify that ϕ_L is indeed a unitary representation of $U(d)$, since the Liouville representation of composition is matrix multiplication and it can easily be verified that $\phi_L(U^\dagger) = \phi_L(U)^\dagger$.

Any representation ϕ of a semisimple group $\mathcal{G}$ [such as $SU(d)$] over a vector space V can be unitarily decomposed into a direct sum of irreducible representations (irreps) $\oplus_l \ \phi_l \otimes \mathbb{1}_{n_l}$, where the l label the irreps and the n_l are the corresponding multiplicities and a rep ϕ over a vector space V is irreducible if there are no nontrivial subspaces of V that are invariant under the action of ϕ. A particularly important irrep for this paper is the trivial irrep ϕ_T such that $\phi_T(g) = 1$ for all $g \in \mathcal{G}$.

In the Liouville representation, vectors $b \in \mathbb{C}^{d^2}$ are in one-to-one correspondence with operators $B \in \mathbb{C}^{d\times d}$, so invariant (vector) subspaces under the Liouville representation can be identified with operator subspaces that are invariant under conjugation in the canonical (i.e., $d \times d$ matrix) representation. In particular, the identity operator $\mathbb{1}$ is invariant under conjugation by any unitary, so $|\mathbb{1})$ is an invariant subspace of the Liouville representation corresponding to a trivial irrep. We now fix $A_1 = \mathbb{1}/\sqrt{d}$ (so that $\langle A_1, A_1\rangle = 1$), so that the Liouville representation of any unitary $\mathcal{U}$ is

$$\phi_L(U) = 1 \oplus \phi_{\rm u}(U), \tag{17}$$

where $\oplus$ denotes the matrix direct sum and we refer to $\phi_{\rm u}(U)$ as the *unital irrep*, which has dimension $d^2 - 1$. Furthermore, any CP channel $\mathcal{E}$ can be written in a corresponding block form as

$$\boldsymbol{\mathcal{E}} = \begin{pmatrix} S(\mathcal{E}) & \boldsymbol{\mathcal{E}}_{\rm sdl} \\ \boldsymbol{\mathcal{E}}_{\rm n} & \boldsymbol{\mathcal{E}}_{\rm u} \end{pmatrix}, \tag{18}$$

where we refer to $\boldsymbol{\mathcal{E}}_{\rm sdl}, \boldsymbol{\mathcal{E}}_{\rm n}$ and $\boldsymbol{\mathcal{E}}_{\rm u}$ as the *state-dependent leakage*, *nonunital* and *unital* blocks respectively. We now show how $\boldsymbol{\mathcal{E}}_{\rm u}$ is related to the definition of the unitarity in equation (4).

Proposition 1. *The unitarity of a channel $\mathcal{E}$ is*

$$u(\mathcal{E}) = \frac{1}{d^2 - 1} \operatorname{Tr} \boldsymbol{\mathcal{E}}_{\rm u}^\dagger \boldsymbol{\mathcal{E}}_{\rm u} \tag{19}$$

Proof. For any operator A, $\operatorname{Tr} A^\dagger A = (A|A)$ and $\boldsymbol{\mathcal{E}}' = P_{\rm u} \boldsymbol{\mathcal{E}} P_{\rm u}$ where $P_{\rm u}$ is the projector onto the unital irrep, so equation (4) can be rewritten as

$$\begin{aligned} u(\mathcal{E}) &= \frac{d}{d-1} \int \mathrm{d}\psi \left(\psi \middle| P_{\rm u} \boldsymbol{\mathcal{E}}^\dagger P_{\rm u} \boldsymbol{\mathcal{E}} P_{\rm u} \middle| \psi \right) \\ &= \frac{d}{d-1} \operatorname{Tr} \boldsymbol{\mathcal{E}}_{\rm u}^\dagger \boldsymbol{\mathcal{E}}_{\rm u} \mathcal{O}, \end{aligned} \tag{20}$$

where $\mathcal{O} = \int \mathrm{d}\psi \, |\psi)(\psi|$, with the slight abuse of notation $\boldsymbol{\mathcal{E}}_{\rm u} = P_{\rm u} \boldsymbol{\mathcal{E}} P_{\rm u}$.

Since $\mathcal{O}$ commutes with the action of the unitary group, Schur's lemma implies that it is a weighted sum of projectors onto the irreps of ϕ_L

$$\mathcal{O} = \lambda_T P_T + \lambda_{\rm u} P_{\rm u}. \tag{21}$$

The projector onto the trivial irrep is $P_T = |A_1)(A_1|$ and so

$$\frac{\operatorname{Tr} P_T \mathcal{O}}{\operatorname{Tr} P_T} = \lambda_T = \int \mathrm{d}\psi \left| \left(A_1 \middle| \psi \right) \right|^2 = \frac{1}{d}. \tag{22}$$

Because $\operatorname{Tr}\mathcal{O} = 1$ from the normalization of the Haar measure, we can solve for $\lambda_{\rm u}$ in the expression

$$\operatorname{Tr} \mathcal{O} = 1 = \frac{1}{d}(1) + \lambda_{\rm u} \left(d^2 - 1 \right) \tag{23}$$

and we find $\lambda_{\rm u} = 1/d(d+1)$. Plugging this in and using $P_{\rm u} P_T = 0$ gives the final result. □

Before we derive the decay curve in equation (14) using the expression for the unitarity from proposition 1, let us first simplify the quantity of interest. The expectation value of Q given that the sequence $\mathbf{j}$ was applied is

$$Q_{\mathbf{j}} = \left(Q \middle| \boldsymbol{\mathcal{U}}_{j_m} \boldsymbol{\mathcal{E}} \ldots \boldsymbol{\mathcal{U}}_{j_2} \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{U}}_{j_1} \middle| \rho \right), \tag{24}$$

where a residual noise term has been absorbed into the experimental state preparation ρ. Noting that

$$Q_{\mathbf{j}}^2 = \left(Q^{\otimes 2} \middle| \boldsymbol{\mathcal{U}}_{j_m}^{\otimes 2} \boldsymbol{\mathcal{E}}^{\otimes 2} \ldots \boldsymbol{\mathcal{U}}_{j_2}^{\otimes 2} \boldsymbol{\mathcal{E}}^{\otimes 2} \boldsymbol{\mathcal{U}}_{j_1}^{\otimes 2} \middle| \rho^{\otimes 2} \right), \tag{25}$$

the expected average of the squares is

$$\begin{aligned} \mathbb{E}_{\mathbf{j}} \left[Q_{\mathbf{j}}^2 \right] &= |\mathcal{G}|^{-m} \sum_{\mathbf{j}} Q_{\mathbf{j}}^2 \\ &= \left(Q^{\otimes 2} \middle| \left(\boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} \boldsymbol{\mathcal{E}}^{\otimes 2} \boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} \right)^{m-1} \middle| \rho^{\otimes 2} \right) \\ &= \left(Q^{\otimes 2} \middle| \boldsymbol{\mathcal{M}}^{m-1} \middle| \rho^{\otimes 2} \right), \end{aligned} \tag{26}$$

where $\boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} = |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \boldsymbol{g}^{\otimes 2}$, we define the averaged operator $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} \boldsymbol{\mathcal{E}}^{\otimes 2} \boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2}$, and we have used the fact that $|\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \phi(g)$ is the projector onto the trivial subreps for any rep ϕ of a group $\mathcal{G}$ so that $\boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} = (\boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2})^2$ [20]. Thus, to derive the fit model we must first identify the trivial irreps of $\mathcal{G}$ in $\phi_L(U)^{\otimes 2}$, since this is where $\boldsymbol{\mathcal{M}}$ is supported.

Proposition 2. *The averaged operator $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2} \boldsymbol{\mathcal{E}}^{\otimes 2} \boldsymbol{\mathcal{U}}_{\rm avg}^{\otimes 2}$ is supported on a two-dimensional subspace spanned by $|\mathbb{1}_{d^2})$ and $|S)$, where S is the SWAP operator.*

Proof. Define $\chi_R(g) = \mathrm{Tr}\, R(g)$ as the character of the rep R. Then we can use Schur's orthogonality relations to count the number of trivial irreps. Let $\langle\langle \chi_R, \chi_{R'} \rangle\rangle = |\mathcal{G}|^{-1} \sum_{g\in\mathcal{G}} \chi_R^*(g)\chi_{R'}(g)$ denote the character inner product for $\mathcal{G}$. From the direct sum structure in equation (17), the number of trivial irreps is

$$\left\langle\!\!\left\langle \chi_1, \chi_{L\otimes L} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \chi_1, \chi_1 + 2\chi_{\mathrm{u}} + \chi_{\mathrm{u}}^2 \right\rangle\!\!\right\rangle = 1 + 0 + \left\langle\!\!\left\langle \chi_1, \chi_{\mathrm{u}}^2 \right\rangle\!\!\right\rangle.$$

Since χ_{u} is real-valued, we have $\langle\langle \chi_1, \chi_{\mathrm{u}}^2 \rangle\rangle = \langle\langle \chi_{\mathrm{u}}, \chi_{\mathrm{u}} \rangle\rangle$. If $\mathcal{G}$ acts irreducibly on the unital block [21], then $\langle\langle \chi_{\mathrm{u}}, \chi_{\mathrm{u}} \rangle\rangle = 1$ and the number of trivial irreps is 2.

The two trivial irreps in $\phi_L^{\otimes 2}$ are spanned by the orthonormal vectors $|B_1)$ and $|B_2)$ where

$$B_1 = \mathbb{1}_{d^2}/d, \quad B_2 = \left(S - \mathbb{1}_{d^2}/d\right)/\sqrt{d^2-1}, \tag{27}$$

and S is the SWAP operator. To check this, note that

$$\left(B_j \middle| \phi_L(U)^{\otimes 2} \middle| B_k\right) = \mathrm{Tr}\left(B_j^\dagger U \otimes U B_k U^\dagger \otimes U^\dagger\right) = \delta_{jk},$$

since both identity and SWAP are invariant under conjugation by $U \otimes U$ and $S^2 = \mathbb{1}_{d^2}$. Since $\phi_L(U)^{\otimes 2}$ is a unitary rep, B_1 and B_2 are the first two elements of a two-qudit orthonormal Schur basis $\{B_j\}$ for $\phi_L^{\otimes 2}$ and so correspond to trivial irreps. Therefore $\boldsymbol{\mathcal{M}}$ is zero except for the 2×2 submatrix supported on $|B_1)$ and $|B_2)$. These vectors have the same span as $|\mathbb{1}_{d^2})$ and $|S)$. □

The next proposition characterizes the averaged operator on the supported subspace.

Proposition 3. *In the invariant* $\mathrm{B_i}$ *basis from equation (27), the averaged operator* $\boldsymbol{\mathcal{M}}$ *has the following matrix elements*

- $\boldsymbol{\mathcal{M}}_{11} = S(\mathcal{E})^2$,
- $\boldsymbol{\mathcal{M}}_{12} = (d^2-1)^{-1/2}\|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2$,
- $\boldsymbol{\mathcal{M}}_{21} = (d^2-1)^{-1/2}\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2$, *and*
- $\boldsymbol{\mathcal{M}}_{22} = (d^2-1)^{-1}\|\boldsymbol{\mathcal{E}}_{\mathrm{u}}\|_F^2 = u(\mathcal{E})$.

Proof. We will establish the matrix elements with respect to $|\mathbb{1}_{d^2})$ and $|S)$, and the claims about the B_i basis will follow by taking appropriate linear combinations.

Because the B_i basis is invariant, we can ignore the average unitary terms in $\boldsymbol{\mathcal{M}}$. We first find that

$$\begin{aligned}\left(\mathbb{1}_{d^2} \middle| \boldsymbol{\mathcal{E}}^{\otimes 2} \middle| \mathbb{1}_{d^2}\right) &= \mathrm{Tr}\, \mathcal{E}^{\otimes 2}\left(\mathbb{1}_{d^2}\right) = \mathrm{Tr}\left(\mathcal{E}\left(\mathbb{1}_d\right)^{\otimes 2}\right) \\ &= \left(\mathrm{Tr}\, \mathcal{E}\left(\mathbb{1}_d\right)\right)^2 = d^2 S(\mathcal{E})^2. \end{aligned} \tag{28}$$

Next we can use the identity $\langle S, A \otimes B\rangle = \mathrm{Tr}\,[S(A \otimes B)] = \mathrm{Tr}(AB)$ and the fact that $\mathcal{E}(A^\dagger) = \mathcal{E}(A)^\dagger$ to find

$$\begin{aligned}\left(S \middle| \boldsymbol{\mathcal{E}}^{\otimes 2} \middle| \mathbb{1}_{d^2}\right) &= \mathrm{Tr}\left[S\mathcal{E}\left(\mathbb{1}_d\right)^{\otimes 2}\right] = \mathrm{Tr}\left[\mathcal{E}\left(\mathbb{1}_d\right)\mathcal{E}\left(\mathbb{1}_d\right)\right] \\ &= \left(\mathcal{E}\left(\tfrac{\mathbb{1}}{d}\right)\middle|\mathcal{E}\left(\tfrac{\mathbb{1}}{d}\right)\right) = \|\boldsymbol{\mathcal{E}}_n\|^2. \end{aligned} \tag{29}$$

The expression for $(\mathbb{1}_{d^2}|\boldsymbol{\mathcal{E}}^{\otimes 2}|S)$ follows similarly using the adjoint channel. Finally, we can use the expansion $S = \sum_k A_k \otimes A_k^\dagger$ for any orthonormal operator basis A_k to obtain

$$\begin{aligned}\left(S \middle| \boldsymbol{\mathcal{E}}^{\otimes 2} \middle| S\right) &= \mathrm{Tr}\left[S\boldsymbol{\mathcal{E}}^{\otimes 2}(S)\right] = \sum_k \mathrm{Tr}\left[S\mathcal{E}\left(A_k\right) \otimes \mathcal{E}\left(A_k^\dagger\right)\right] \\ &= \sum_k \mathrm{Tr}\left[\mathcal{E}\left(A_k\right)^\dagger \mathcal{E}\left(A_k\right)\right] = \|\boldsymbol{\mathcal{E}}\|_F^2, \end{aligned} \tag{30}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The values of the matrix elements are then established by using the form of equation (18) and the definition of the B_i basis from equation (27) and taking various linear combinations. □

The final step in deriving the fit model is to analyze the eigenvalues and eigenvectors of the averaged operator.

Proposition 4. *The averaged operator $\mathcal{M}$ has two distinct nontrivial eigenvectors.*

Proof. Since the averaged operator vanishes almost everywhere, we only need to consider the 2×2 submatrix derived above. The nontrivial eigenvalues are

$$\lambda_{\pm} = \frac{1}{2}\mathcal{M}_{11} + \frac{1}{2}\mathcal{M}_{22} \pm \frac{1}{2}\sqrt{\left[\mathcal{M}_{11} - \mathcal{M}_{22}\right]^2 + 4\mathcal{M}_{12}\mathcal{M}_{21}}. \tag{31}$$

This spectrum is degenerate precisely when the terms under the square root both vanish (since both terms are nonnegative). Whenever the spectrum is nondegenerate, there are trivially two distinct eigenvectors, so we only need to deal with the degenerate case.

We will break the analysis for the degenerate spectrum into two nontrivial cases, $\mathcal{M}_{11} = \mathcal{M}_{22}$ and either $\mathcal{M}_{12} = 0$ or $\mathcal{M}_{21} = 0$, exclusive. There are also two trivial cases: when $\mathcal{M}_{12} = \mathcal{M}_{21} = 0$, the matrix $\mathcal{M}$ is already diagonal and we are done. We ignore the pathological case when $\mathcal{M}_{11} = 0$, since this corresponds physically to a state that is never observable. In both nontrivial cases, we will make use of the two-qudit state $\Pi_a = \dfrac{\mathbb{1} - S}{d(d-1)}$, the maximally mixed state on the antisymmetric subspace. Expanding this state in the B_i basis gives

$$\left|\Pi_a\right) = \pi_1 \left|B_1\right) + \pi_2 \left|B_2\right) = \frac{1}{d}\left|B_1\right) - \frac{\sqrt{d^2-1}}{d(d-1)}\left|B_2\right).$$

The key feature of this state is that $\pi_2 < 0$.

Case 1: $\mathcal{M}_{12} = 0$. In this case

$$\mathcal{M} = \begin{pmatrix} \lambda & 0 \\ y & \lambda \end{pmatrix}, \tag{32}$$

where $\lambda > 0$ and $y \geqslant 0$. Taking the mth power gives

$$\mathcal{M}^m = \lambda^{m-1}\begin{pmatrix} \lambda & 0 \\ my & \lambda \end{pmatrix}. \tag{33}$$

If we perform the measurement $\{\Pi_a, \mathbb{1} - \Pi_a\}$ on a system prepared in the state $\mathbb{1}_{d^2}/d^2$ which evolves under $\mathcal{M}^m$, then the probability of observing the outcome Π_a is

$$\left(\Pi_a\middle|\mathcal{M}^m\middle|\mathbb{1}_{d^2}/d^2\right) = \frac{\lambda^{m-1}}{d}\left(\lambda\pi_1 + my\pi_2\right). \tag{34}$$

Since $\lambda,\ \pi_1 > 0, y \geqslant 0$, and $\pi_2 < 0$, in order for this to be a probability for all m, we require $y = 0$ and so $\mathcal{M}$ is actually diagonal.

Case 2: $\mathcal{M}_{21} = 0$. In this case,

$$\mathcal{M} = \begin{pmatrix} \lambda & y \\ 0 & \lambda \end{pmatrix}, \tag{35}$$

where $\lambda > 0$ and $y \geqslant 0$. Taking the mth power gives

$$\mathcal{M}^m = \lambda^{m-1}\begin{pmatrix} \lambda & my \\ 0 & \lambda \end{pmatrix}. \tag{36}$$

Therefore the probability of detecting the system (i.e., measuring $\mathbb{1}_d^2$) when a system is prepared in the state Π_a and evolves under $\mathcal{M}^m$ is

$$\left(\mathbb{1}_{d^2}\middle|\mathcal{M}^m\middle|\Pi_a\right) = \lambda^{m-1}\left(\lambda\pi_1 + my\pi_2\right). \tag{37}$$

Again since $\lambda,\ \pi_1 > 0, y \geqslant 0$, and $\pi_2 < 0$, for this to be a valid probability for all m, we require $y = 0$ and so $\mathcal{M}$ is actually diagonal. □

We now have all the ingredients to derive the fit models of equations (8) and (14).

Theorem 5. *For time- and gate-independent noise, the expected value $\mathbb{E}_{\mathbf{j}}[Q_{\mathbf{j}}^2]$ obeys the decay equation*

$$\mathbb{E}_{\mathbf{j}}\left[Q_{\mathbf{j}}^2\right] = A + Bu(\mathcal{E})^{m-1}$$

for trace-preserving noise, and for trace-decreasing noise it obeys

$$\mathbb{E}_{\mathrm{j}}\left[Q_{\mathrm{j}}^2\right] = A\lambda_+^{m-1} + B\lambda_-^{m-1},$$

where $\lambda_\pm$ are given by equation (31), $\lambda_+ + \lambda_- = S(\mathcal{E})^2 + u(\mathcal{E})$, and the constants A and B depend only on state preparation and measurement errors and the unitary that diagonalizes $\mathcal{M}$.

Proof. Proposition 4 establishes that the matrix $\mathcal{M}$ is diagonalizable by a similarity transform with eigenvalues given by equation (31). From equation (26), we can diagonalize $\mathcal{M}$ and absorb the similarity transform into $|\rho^{\otimes 2})$ and $(Q^{\otimes 2}|$ as SPAM, yielding

$$\begin{aligned}\mathbb{E}_{\mathrm{j}}\left[Q_{\mathrm{j}}^2\right] &= \left(Q^{\otimes 2}\middle|\mathcal{M}^{m-1}\middle|\rho^{\otimes 2}\right)\\ &= A\lambda_+^{m-1} + B\lambda_-^{m-1}.\end{aligned} \tag{38}$$

Trace-preserving noise is a special case of this, since if $\mathcal{E}$ is TP, then by proposition 3 we have $\lambda_+ = S(\mathcal{E})^2 = 1$ and so $\lambda_- = u(\mathcal{E})$. □

We note that the unitary that diagonalizes $\mathcal{M}$ will in general depend on the noise channel, and hence will depend on u. We conflate this dependence with the SPAM errors in our fit model, as the diagonalization of $\mathcal{M}$ does not depend on the sequence length m. Neglecting the dependence on u thus results in a model that is correct, but is slightly less sensitive to u than is optimal.

We are now equipped to formalize the observation made in section 2 that the optimal observable for trace-preserving noise is an operator proportional to B_2. This follows from noting that such operators overlap fully with the component of $\mathcal{M}$ that give rise to the exponential term, as given by equation (30).

5. Properties of the unitarity

We now prove some properties of the unitarity for CPTP channels that make it a practical quantification of the coherence of a channel. We begin by proving that the unitarity and the average incoherent survival probability can be used to bound the nonunital and state-dependent leakage terms which are subtracted off in the definition of unitarity in equation (4).

Proposition 6. *For any channel $\mathcal{E}$,*

$$\max\left\{\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2, \|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2\right\} \leqslant \tfrac{1}{2}\left(d^2-1\right)\left[S(\mathcal{E})^2 - u(\mathcal{E})\right]. \tag{39}$$

If $\mathcal{E}$ is trace-preserving, then $\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2 \leqslant (d-1)[1-u(\mathcal{E})]$.

Proof. Consider the maximally mixed states on the symmetric and antisymmetric subspaces, $\Pi_s = \frac{\mathbb{1}+S}{d(d+1)}$ and $\Pi_a = \frac{\mathbb{1}-S}{d(d-1)}$ respectively, and let E_s and E_a be the respective projectors onto these spaces. Expanding these states in the B_i basis gives

$$\begin{aligned}\left|\Pi_s\right) &= \frac{1}{d}\left|B_1\right) + \frac{\sqrt{d^2-1}}{d(d+1)}\left|B_2\right)\\ \left|\Pi_a\right) &= \frac{1}{d}\left|B_1\right) - \frac{\sqrt{d^2-1}}{d(d-1)}\left|B_2\right).\end{aligned} \tag{40}$$

Preparing the state Π_s (Π_a), evolving under $\mathcal{M}$ and then measuring the POVM $\{E_a, \mathbb{1}-E_a\}$ ($\{E_s, \mathbb{1}-E_s\}$) produces the outcomes E_a (E_s) with probabilities

$$\begin{aligned}p_{as} &= \left(E_a\middle|\mathcal{M}\middle|\Pi_s\right)\\ &= \frac{d-1}{2d}\left(S(\mathcal{E})^2 - u(\mathcal{E}) - \frac{\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2}{d-1} + \frac{\|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2}{d+1}\right)\end{aligned} \tag{41}$$

$$\begin{aligned}p_{sa} &= \left(E_s\middle|\mathcal{M}\middle|\Pi_a\right)\\ &= \frac{d+1}{2d}\left(S(\mathcal{E})^2 - u(\mathcal{E}) + \frac{\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2}{d+1} - \frac{\|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2}{d-1}\right)\end{aligned} \tag{42}$$

respectively, where we have used proposition 3. Since both these expressions are probabilities we have $0 \leqslant p_{as} \leqslant 1$ and $0 \leqslant p_{sa} \leqslant 1$. Taking appropriate linear combinations of these two inequalities will cancel the

dependence on either $\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2$ or $\|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2$, isolating the other variable. Simplifying the resulting expressions gives the bound equation (39) for both quantities individually, hence the maximum holds as well.

Furthermore, if the noise is trace-preserving, then $\|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2 = 0$ and $S(\mathcal{E}) = 1$, so $p_{as} \geqslant 0$ gives $\|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2 \leqslant (d-1)[1-u(\mathcal{E})]$ for trace-preserving noise. □

We now prove that $u(\mathcal{E}) = 1$ if and only if $\mathcal{E}$ is unitary and that $u(\mathcal{E})$ is invariant under composition with unitaries.

Proposition 7. *For any channel $\mathcal{E}$, $u(\mathcal{E}) \leqslant 1$ with equality if and only if $\mathcal{E}$ is unitary. Furthermore, the unitarity satisfies $u(\mathcal{V} \circ \mathcal{E} \circ \mathcal{U}) = u(\mathcal{E})$ for any unitaries U, $V \in U(d)$.*

Proof. The unitary invariance $u(\mathcal{V} \circ \mathcal{E}\circ\mathcal{U}) = u(\mathcal{E})$ follows immediately from the invariance of the trace under cyclic permutations.

Since the norms of vectors are always nonnegative, $u(\mathcal{E}) = 1$ only if $\mathcal{E}$ is trace-preserving and unital by equation (39), in which case the adjoint channel $\mathcal{E}^\dagger$ is also a channel [22] and so the eigenvalues of $\boldsymbol{\mathcal{E}}^\dagger\boldsymbol{\mathcal{E}}$ (i.e., the singular values of $\boldsymbol{\mathcal{E}}$) are all bounded by one [23]. Therefore $u(\mathcal{E}) = 1$ only if $\mathcal{E}$ is unital and all the eigenvalues of $\boldsymbol{\mathcal{E}}$ have unit modulus and consequently if $|\det \boldsymbol{\mathcal{E}}| = 1$. However, the only channels with $|\det \boldsymbol{\mathcal{E}}| = 1$ are unitary channels [24]. Since $u(\mathcal{E})$ is unitarily invariant and $u(\mathcal{I}) = 1$, $u(\mathcal{E}) = 1$ if and only if $\mathcal{E}$ is unitary, as claimed. □

We now show that the unitarity can be used with the average gate infidelity to quantify the intermediate regime between incoherent and unitary errors. It is useful to define a notion of average gate infidelity that has been optimized to remove unitary noise. First recall the definition of average gate infidelity,

$$r(\mathcal{E}) = 1 - \int \mathrm{d}\psi\ \mathrm{Tr}[\psi\mathcal{E}(\psi)], \tag{43}$$

Then for any CPTP channel $\mathcal{E}$, define

$$R(\mathcal{E}) = \min_{U\in U(d)} r(\mathcal{E}\circ\mathcal{U}). \tag{44}$$

where the equivalance follows from the Haar invariance in equation (43). This quantity can be thought of as the best average gate infidelity that is achievable with perfect unitary control. For example, if $\mathcal{E}$ is a unitary channel, then $R(\mathcal{E}) = 0$.

Proposition 8. *For any CPTP channel $\mathcal{E}$ with average gate infidelity $r = r(\mathcal{E})$ to the identity and $R = R(\mathcal{E})$ as above, then the following inequalities hold*

$$u(\mathcal{E}) \geqslant [1 - dR/(d-1)]^2 \geqslant [1 - dr/(d-1)]^2. \tag{45}$$

The chain of inequalities is saturated if and only if $\mathcal{E}$ has a unital block $\boldsymbol{\mathcal{E}}_{\mathrm{u}}$ that is a diagonal scalar matrix.

Proof. Any channel with infidelity r to the identity can be written as $\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{I}} - r\Delta$ where the diagonal entries of Δ are nonnegative and $\mathrm{Tr}\,\Delta = d(d+1)$. We then have

$$\|\boldsymbol{\mathcal{E}}_{\mathrm{u}}\|_2^2 \geqslant \sum_{k=2}^{d^2}\left(1 - \Delta_{kk}r\right)^2 = d^2 - 1 - 2r\ \mathrm{Tr}\,\Delta + \sum_{k=2}^{d^2}\Delta_{kk}^2 r^2,$$

with equality if and only if $\boldsymbol{\mathcal{E}}_{\mathrm{u}}$ is diagonal. The term $\sum_k \Delta_{kk}^2$ is uniquely minimized for nonnegative Δ_{kk} subject to the constraint $\mathrm{Tr}\,\Delta = d(d+1)$ by setting $\Delta_{kk} = d/(d-1)$ (that is, by setting all the diagonal entries to be equal). This proves the weaker inequality bounding $u(\mathcal{E})$ in terms of r. To get the stronger inequality in terms of $R(\mathcal{E})$, we use the unitary invariance proven in proposition 7 and optimize the inequality over all unitary channels. □

We note that the first inequality in equation (45) is saturated at 1 when the noise channel is unitary, and the chain of inequalities is saturated for depolarizing noise, or depolarizing noise composed with amplitude damping.

An immediate corollary of proposition 8 is that the unitarity can be used to put a lower bound on the best possible average infidelity in the presence of perfect unitary control. Rearranging equation (45), we find a lower bound

$$\frac{d-1}{d}\left(1 - \sqrt{u(\mathcal{E})}\right) \leqslant R(\mathcal{E}) \leqslant r(\mathcal{E}). \tag{46}$$

The unitarity is also closely related to the purity of the Jamiłłkowski state associated to the noise channel.

Proposition 9. *The unitarity is related to the purity of the Jamiołkowski state by*

$$d^2 \operatorname{Tr}\left[J(\mathcal{E})^\dagger J(\mathcal{E})\right] = S(\mathcal{E})^2 + \|\boldsymbol{\mathcal{E}}_{\mathrm{sdl}}\|^2 + \|\boldsymbol{\mathcal{E}}_{\mathrm{n}}\|^2 + \left(d^2 - 1\right)u(\mathcal{E}), \tag{47}$$

where $J(\mathcal{E}) = (\mathcal{E} \otimes \mathcal{I})[\Phi]$ *and* $\Phi = \frac{1}{d}\sum_{j,k}|jj\rangle\langle kk|$.

Proof. We begin with an alternate representation of the maximally entangled state $\Phi = d^{-1}\sum_k A_k \otimes A_k$. By cycling the adjoint channel in the trace, the purity of $J(\mathcal{E})$ becomes

$$\begin{aligned}
\operatorname{Tr}\left[J(\mathcal{E})^\dagger J(\mathcal{E})\right] &= \operatorname{Tr}\left[\Phi\left(\mathcal{E}^\dagger\mathcal{E} \otimes \mathcal{I}\right)[\Phi]\right] \\
&= \left(\Phi\middle|\boldsymbol{\mathcal{E}}^\dagger\boldsymbol{\mathcal{E}} \otimes \boldsymbol{\mathcal{I}}\middle|\Phi\right) \\
&= \frac{1}{d^2}\sum_{j,k}\left(A_j \otimes A_j\middle|\boldsymbol{\mathcal{E}}^\dagger\boldsymbol{\mathcal{E}} \otimes \boldsymbol{\mathcal{I}}\middle|A_k \otimes A_k\right) \\
&= \frac{1}{d^2}\sum_{j,k}\left(A_j\middle|\boldsymbol{\mathcal{E}}^\dagger\boldsymbol{\mathcal{E}}\middle|A_k\right)\left(A_j\middle|A_k\right).
\end{aligned} \tag{48}$$

Since the A_k are a trace orthonormal basis, the last line simplifies to

$$\operatorname{Tr}\left[J(\mathcal{E})^\dagger J(\mathcal{E})\right] = \tfrac{1}{d^2}\operatorname{Tr}\,\boldsymbol{\mathcal{E}}^\dagger\boldsymbol{\mathcal{E}}. \tag{49}$$

Comparing this expression to the decomposition in equation (18) and using proposition 1 completes the proof. □

Finally, we give a simple example that shows that the unitarity is *not* a monotone, in the sense that it can oscillate under composition of channels. Consider the two (nearly) dual qubit channels,

$$\mathcal{E}_0(\rho) = \operatorname{Tr}(\rho)|0\rangle\langle 0| \quad \text{and} \quad \frac{1}{2}\mathcal{E}_0^\dagger(\rho) = \langle 0|\,\rho\,|0\rangle\frac{\mathbb{1}}{2}.$$

Then the unitarity of both $\mathcal{E}_0$ and $\frac{1}{2}\mathcal{E}_0^\dagger$ is zero, while the unitarity of the composed channel $\frac{1}{2}\boldsymbol{\mathcal{E}}_0\boldsymbol{\mathcal{E}}_0^\dagger$ is 1/12.

We note that for some restricted classes of channels the unitarity is indeed a monotone. For example, a trivial application of von Neumann's trace inequality shows that if the singular values of the unital block are all less than or equal to 1 (which holds for all qubit channels and all unital channels), then it is a monotone for trace-preserving channels.

6. Conclusion

In this paper, we have shown that the coherence of a noisy process can be quantified by the unitarity, which corresponds to the change in the purity (with the identity components subtracted off) averaged over pure states. We have presented a protocol for efficiently estimating the unitarity of the average noise in the implementation of a unitary 2-design.

We have also proven that the unitarity is 1 if and only if the noise source is unitary and provided a tight lower bound for the unitarity in terms of the infidelity (which can be estimated using randomized benchmarking [8]). This allows the intermediate regime between fully incoherent and unitary errors to be quantified, potentially allowing for improved bounds on the worst-case error. We have also shown that the unitarity provides a lower bound on the best achievable gate infidelity assuming perfect unitary control.

Our present results also have direct implications for the loss protocol when applied to a unitary 2-design, since the variance over random sequences of fixed length for the protocol in [16] is

$$\mathbb{V}_{\mathrm{j}}\left(Q_{\mathrm{j}}\right) = \mathbb{E}_{\mathrm{j}}\left(Q_{\mathrm{j}}^2\right) - \left[\mathbb{E}_{\mathrm{j}}\left(Q_{\mathrm{j}}\right)\right]^2, \tag{50}$$

which decays faster with m for fixed $S(\mathcal{E})$ if the unitarity is smaller (and hence the two decay rates in the fit curve for determining the unitarity, $\lambda_\pm$, are smaller). A lower variance over sequences allows a more precise estimation of the average incoherent survival probability for a fixed number of experiments. Similar implications may also hold for standard randomized benchmarking since $u(\mathcal{E})$ can easily be seen to be one of the eigenvalues of the averaged operator in [13] that determines the variance and is precisely the eigenvalue that determines the asymptotic variance. However, in order to establish a concrete bound, it would have to be shown that $u(\mathcal{E})$ is in fact the largest eigenvalue.

There are four important open problems raised by this work. First, while the unitarity is a monotone for unital noise, it is not a monotone for trace-decreasing noise. We leave open the problem of finding necessary and sufficient conditions for when $u(\mathcal{E})$ is a monotone, or finding other quantities that are monotonic in general.

Second, our protocol characterizes the unitarity of the average noise, but does not characterize the unitarity of the errors in the individual gate. While a variant of interleaved randomized benchmarking [25] should hold for the current protocol, obtaining reasonable bounds on the unitarity of the individual error is an open problem.

Third, the signal for our protocol is substantially improved by the purity measurement, but the method of performing the purity measurement via measuring Pauli operators is not scalable beyond a handful of qubits because of the exponential size of the Pauli group on n qubits. Moreover, measuring any single Pauli operator will in general give a small signal as the number of qubits grows, since we do not perform an inversion step. Directly using the SWAP operation on two copies of the system running in parallel is a mathematical solution, but the extra resources required to implement this might be prohibitive and an analysis of the role of crosstalk and correlations would be required to justify this idea. Thus, identifying efficient measurements that give a good signal on multi-qubit systems remains an open problem.

Finally, a pressing open problem identified in this paper is to obtain an improved bound on the worst-case error in terms of both the infidelity and the unitarity. Such a bound would substantially reduce the effort required to certify that an experimental implementation is near (or below) the threshold for fault-tolerant quantum computation.

Acknowledgments

We thank Jay Gambetta for pointing out that estimating purity would reduce variance. This work was supported by the ARC via EQuS project number CE11001013, by IARPA via the MQCO program, and by the US Army Research Office grant numbers W911NF-14-1-0098 and W911NF-14-1-0103. STF also acknowledges support from an ARC Future Fellowship FT130101744.

References

[1] Chuang I L and Nielsen M A 1997 Prescription for experimental determination of the dynamics of a quantum black box *J. Mod. Opt.* **44** 2455
[2] Poyatos J, Cirac J and Zoller P 1997 Complete characterization of a quantum process: the two-bit quantum gate *Phys. Rev. Lett.* **78** 390
[3] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Córcoles A D, Johnson B R, Ryan C A and Steffen M 2013 Self-consistent quantum process tomography *Phys. Rev.* A **87** 062119
[4] Emerson J, Alicki R and Życzkowski K 2005 Scalable noise estimation with random unitary operators *J. Opt. B: Quantum Semiclass. Opt.* **7** S347
[5] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 Randomized benchmarking of quantum gates *Phys. Rev.* A **77** 012307
[6] Dankert C, Cleve R, Emerson J and Livine E 2009 Exact and approximate unitary 2-designs and their application to fidelity estimation *Phys. Rev.* A **80** 012304
[7] López C C, Bendersky A, Paz J P and Cory D G 2010 Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies *Phys. Rev.* A **81** 062113
[8] Magesan E, Gambetta J M and Emerson J 2011 Scalable and robust randomized benchmarking of quantum processes *Phys. Rev. Lett.* **106** 180504
[9] Magesan E, Gambetta J M and Emerson J 2012 Characterizing quantum gates via randomized benchmarking *Phys. Rev.* A **85** 042311
[10] Flammia S T and Yi-Kai L 2011 Direct fidelity estimation from few pauli measurements *Phys. Rev. Lett.* **106** 230501
[11] da Silva M P, Landon-Cardinal O and Poulin D 2011 Practical characterization of quantum devices without tomography *Phys. Rev. Lett.* **107** 210404
[12] Kitaev A 1997 Quantum computations: algorithms and error correction *Russ. Math. Surv.* **52** 1191–249
[13] Wallman J J and Flammia S T 2014 Randomized benchmarking with confidence *New J. Phys.* **16** 103032
[14] Sanders Y R, Wallman J J and Sanders B C 2015 Bounding quantum gate error rate based on reported gate fidelity (arXiv:1501.04932v2)
[15] O'Malley P J J *et al* 2015 Qubit metrology of ultralow phase noise using randomized benchmarking *Phys. Rev. Appl.* **3** 044009
[16] Wallman J J, Barnhill M and Emerson J 2015 Robust characterization of loss rates *Phys. Rev. Lett.* **115** 060501
[17] Baumgratz T, Cramer M and Plenio M B 2014 Quantifying coherence *Phys. Rev. Lett.* **113** 140401
[18] Bruzda W, Cappellini V, Hans-Jürgen S and Życzkowski K 2009 Random quantum operations *Phys. Lett.* A **373** 320–4
[19] Johansson J R, Nation P D and Nori F 2013 QuTiP 2: a Python framework for the dynamics of open quantum systems *Comput. Phys. Commun.* **184** 1234–40
[20] Goodman R and Wallach N R 2009 *Symmetry, Representations, and Invariants* (*Graduate Texts in Mathematics*) (Berlin: Springer)
[21] Gross D, Audenaert K and Eisert J 2007 Evenly distributed unitaries: On the structure of unitary designs *J. Math. Phys.* **48** 052104
[22] Watrous J 2011 *CS 766/QIC 820 Theory of Quantum Information* (*Fall 2011*)
[23] Evans D E and Hoegh-Krohn R 1978 Spectral properties of positive maps on C^*-algebras *J. London Math. Soc.* **17** 345–55
[24] Wolf M M and Cirac J I 2008 Dividing quantum channels *Commun. Math. Phys.* **279** 147–68
[25] Magesan E *et al* 2012 Efficient measurement of quantum gate error by interleaved randomized benchmarking *Phys. Rev. Lett.* **109** 080505