

# Compressed Sensing Measurement of Long-Range Correlated Noise

Alireza Seif,<sup>1,2,3</sup> Mohammad Hafezi,<sup>1,4</sup> and Yi-Kai Liu<sup>2,5</sup>

<sup>1</sup>*Department of Physics and the Joint Quantum Institute (JQI),  
University of Maryland, College Park, MD 20742*

<sup>2</sup>*Joint Center for Quantum Information and Computer Science (QuICS),  
University of Maryland, College Park, MD 20742*

<sup>3</sup>*Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637*

<sup>4</sup>*Department of Electrical and Computer Engineering,  
University of Maryland, College Park, MD 20742*

<sup>5</sup>*Applied and Computational Mathematics Division,  
National Institute of Standards and Technology, Gaithersburg, MD 20899*

(Dated: May 27, 2021)

Long-range correlated errors can severely impact the performance of NISQ (noisy intermediate-scale quantum) devices, and fault-tolerant quantum computation. Characterizing these errors is important for improving the performance of these devices, via calibration and error correction, and to ensure correct interpretation of the results. We propose a compressed sensing method for detecting two-qubit correlated dephasing errors, assuming only that the correlations are *sparse* (i.e., at most  $s$  pairs of qubits have correlated errors, where  $s \ll n(n-1)/2$ , and  $n$  is the total number of qubits). In particular, our method can detect *long-range* correlations between any two qubits in the system (i.e., the correlations are not restricted to be geometrically local).

Our method is highly scalable: it requires as few as  $m = O(s \log n)$  measurement settings, and efficient classical postprocessing based on convex optimization. In addition, when  $m = O(s \log^4 n)$ , our method is highly robust to noise, and has sample complexity  $O(\max(n, s)^2 \log^4(n))$ , which can be compared to conventional methods that have sample complexity  $O(n^3)$ . Thus, our method is advantageous when the correlations are sufficiently sparse, that is, when  $s \leq O(n^{3/2}/\log^2 n)$ . Our method also performs well in numerical simulations on small system sizes, and has some resistance to state-preparation-and-measurement (SPAM) errors. The key ingredient in our method is a new type of compressed sensing measurement, which works by preparing entangled Greenberger-Horne-Zeilinger states (GHZ states) on random subsets of qubits, and measuring their decay rates with high precision.

## I. INTRODUCTION

The development of noisy intermediate-scale quantum information processors (NISQ devices) has the potential to advance many areas of computational science [1]. An important problem is the characterization of noise processes in these devices, in order to improve their performance (via calibration and error correction), and to ensure correct interpretation of the results [2]. The challenge here is to characterize all of the noise processes that are likely to occur in practice, using some experimental procedure that is efficient and can scale up to large numbers of qubits.

Compressed sensing [3] offers one approach to solving this problem. Here one uses specially-designed measurements (and classical postprocessing) to learn an unknown signal that has some prescribed structure. For example, the unknown signal can be a sparse vector or a low-rank matrix, the measurements can consist of random projections sampled from various distributions, and the classical

postprocessing can consist of solving a convex optimization problem (e.g., minimizing the  $\ell_1$  or trace norm), using efficient algorithms. This approach has been used in several previous works on quantum state and process tomography, and estimation of Hamiltonians and Lindbladians [4–8].

From a theoretical perspective, one of the main challenges in this line of work is to design *measurements* that have the mathematical properties needed for compressed sensing, and can be implemented efficiently on a quantum device. There has been a substantial amount of work in this area, which can be broadly classified into two approaches: “sparsity-based” and “low-rank” compressed sensing. For compressed sensing of “low-rank” objects (e.g., low-rank density matrices and quantum processes), there seem to be a few natural choices for measurements, including random Pauli measurements, and fidelities with random Clifford operations [4, 8, 9]. For compressed sensing of “sparse” objects (e.g., sparse Hamiltonians, Lindbladians, or Pauli channels), however, the situation is more com-

licated, as a larger number of different measurement schemes and classical postprocessing methods have been proposed, and the optimal type of measurement seems to depend on the situation at hand [5, 6, 10]. This complicated state of affairs can be explained in part because “sparsity” occurs in a wider variety of situations than “low-rankness.”

In this paper, we extend the theory of “sparsity-based” quantum compressed sensing, and apply it to a physical problem that is relevant to the development of NISQ devices: detecting long-range correlated dephasing errors. We use a simple model of correlated dephasing, which is described by a Markovian master equation:

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = \sum_{j,k=1}^n c_{jk} (Z_k \rho Z_j - \frac{1}{2} \{Z_j Z_k, \rho\}). \quad (\text{I.1})$$

Here the system consists of  $n$  qubits, and  $Z_j$  and  $Z_k$  are Pauli  $\sigma_z$  operators that act on the  $j$ 'th and  $k$ 'th qubits, respectively. The noise is then completely described by the correlation matrix  $C = (c_{jk}) \in \mathbb{R}^{n \times n}$ , see Fig. I.1(a) and (b). (We will also consider generalizations of this model with complex coefficients  $c_{jk}$ , and an additional environment-induced Hamiltonian.)

Here, the diagonal elements  $c_{jj}$  show the rates at which single qubits dephase, and the off-diagonal elements  $c_{jk}$  show the rates at which pairs of qubits undergo correlated dephasing. Typically, the diagonal of  $C$  will be nonzero, while the off-diagonal part may be dense or sparse, depending on the degree of connectivity between the qubits and their environment.

This master equation describes a number of physically plausible scenarios, such as spin-1/2 particles coupling to a shared bosonic bath [11] (see Appendix A). It has also been studied as an example of how collective decoherence can affect physical implementations of quantum computation [12–15] and quantum sensing [16, 17].

This model of correlated dephasing is quite different from other models of crosstalk that are based on quantum circuits or Pauli channels [10, 18–20]. Roughly speaking, our model describes crosstalk that arises from the physical coupling of the qubits to their shared environment. This has a different character from crosstalk that arises from performing imperfect two-qubit gates, or correlated errors that result when the physical noise processes are symmetrized by applying quantum operations such as Pauli twirls. Nonetheless, there are intriguing parallels between our results, and some of these other works, particularly on the estimation of sparse Pauli

channels [10]. We will discuss this later in this section.

In this paper, we show how our model of correlated dephasing can be learned efficiently when the off-diagonal part of the correlation matrix  $C$  is *sparse*. We assume that  $C$  has at most  $s \ll n(n-1)/2$  nonzero entries above the diagonal, but these entries may be distributed arbitrarily; in particular, long-range correlated errors are allowed. (Note that physical constraints imply that  $C$  is positive semidefinite, hence  $C = C^\dagger$  [21].)

This model is applicable in a number of scenarios, including experimental NISQ devices, which are often engineered to have long-range interactions, in order to perform quantum computations more efficiently [22]; the execution of quantum circuits on distributed quantum computers, where long-range correlations are generated when qubits are moved or teleported from one location to another [23]; and quantum sensor arrays, where a detection event at a location  $(j, k)$  in the array may be registered as a pairwise correlation between a qubit that is coupled to row  $j$  and a qubit that is coupled to column  $k$  in the array.

Our main technical contribution is a new method for performing compressed sensing of the coupling matrix  $C$ . At a high level, our method works by performing  $m = O(s \log n)$  or  $m = O(s \log^4 n)$  random linear measurements of the coupling matrix  $C$ , where each measurement can be understood as a generalized Ramsey measurement [24, 25] (Fig. I.1(c) and (d)). At an abstract level, each measurement has the following form: choose two vectors  $\mathbf{a}, \mathbf{b} \in \{0, 1\}^n$  uniformly at random, and estimate the quantity  $\mathbf{r}^T C \mathbf{r}$ , where  $\mathbf{r} = \mathbf{b} - \mathbf{a} \in \{1, 0, -1\}^n$ .

These kinds of measurements can be realized experimentally, using techniques from noise spectroscopy and quantum sensing [16, 17, 25]: prepare an  $n$ -qubit state  $|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle) \in (\mathbb{C}^2)^{\otimes n}$  (assuming  $\mathbf{a} \neq \mathbf{b}$ ), and allow it to evolve according to equation (I.1) for some time  $t$  to get a state  $\rho(t)$ . A straightforward calculation shows that  $\rho(t)$  has the form

$$\rho(t) = \frac{1}{2} \left( |\mathbf{a}\rangle\langle\mathbf{a}| + e^{-\Gamma_{\mathbf{ab}} t} |\mathbf{a}\rangle\langle\mathbf{b}| + e^{-\Gamma_{\mathbf{ab}} t} |\mathbf{b}\rangle\langle\mathbf{a}| + |\mathbf{b}\rangle\langle\mathbf{b}| \right), \quad (\text{I.2})$$

where the off-diagonal elements decay exponentially at a rate  $\Gamma_{\mathbf{ab}} = 2\mathbf{r}^T C \mathbf{r}$ . One can estimate  $\Gamma_{\mathbf{ab}}$  from experiments, which then gives us the desired quantity  $\mathbf{r}^T C \mathbf{r}$ .

This data can be viewed as an estimate of the vector  $\Phi(C)$  returned by the “measurement operator”

$$\Phi : C \mapsto [(\mathbf{r}^{(1)})^T C \mathbf{r}^{(1)}, \dots, (\mathbf{r}^{(m)})^T C \mathbf{r}^{(m)}], \quad (\text{I.3})$$

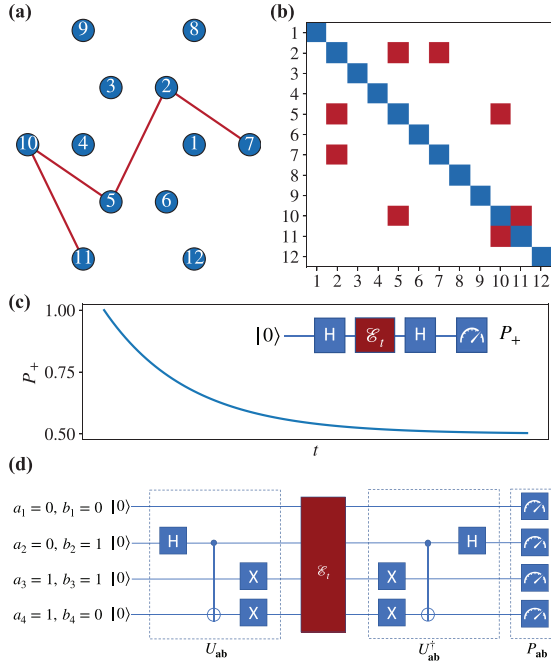


FIG. I.1. Illustration of the noise model and the protocol. (a) The qubits experience correlated Markovian dephasing. The red lines correspond to non-zero  $c_{ij}$ , indicating correlated noise affecting the pairs of qubits connected by those lines. (b) The  $C$  matrix corresponding to the correlation graph in panel (a). The diagonal elements correspond to single qubit dephasing whereas the off-diagonal elements indicate correlated dephasing noise. (c) Single qubit Ramsey spectroscopy. The plot shows the overlap  $P_+$ , which decays exponentially (towards  $1/2$ ) with decay rate  $\Gamma$ , as a function of time  $t$ . The inset shows the Ramsey protocol, where a superposition of qubit states is prepared with the first Hadamard gate  $H$ , the system undergoes dephasing (represented by the noise channel  $\mathcal{E}_t$ ) for time  $t$ , and the second  $H$  followed by a measurement in the computational basis measures the overlap  $P_+$ . (d) The generalized measurement protocol involves generating vectors  $\mathbf{a}$  and  $\mathbf{b}$  whose elements are randomly chosen from  $\{0, 1\}$ . The operation  $U_{\mathbf{ab}}$  prepares the state  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ . The system evolves under dephasing noise for time  $t$ , represented by  $\mathcal{E}_t$ . Finally, we apply  $U_{\mathbf{ab}}^\dagger$  and perform a computational basis measurement. The probability of obtaining the outcome  $0$ ,  $P_{\mathbf{ab}}$ , decays exponentially (towards  $1/2$ ) as  $t$  increases. By measuring the decay rate  $\Gamma_{\mathbf{ab}}$  for various  $\mathbf{a}$ 's and  $\mathbf{b}$ 's we can recover  $C$ .

where  $\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(m)}$  are random vectors sampled independently from the same distribution as  $\mathbf{r}$ . Given a (noisy) estimate of  $\Phi(C)$ , the matrix  $C$  can then be reconstructed (up to some small error) by using techniques from compressed sensing, e.g., by solving

a convex optimization problem such as constrained  $\ell_1$ -minimization or  $\ell_1$ -regularized least-squares regression [26]. We describe the complete method in Sections II and III.

Numerical simulations show that our method performs well, and readily scales up to 128 qubits or more (see Fig. I.2 and Section IV). But the reasons for this success are not at all obvious, because our method is substantially different from previous work on quantum compressed sensing. In particular, the linear measurement  $C \mapsto \mathbf{r}^T C \mathbf{r}$  has an unusual form: it is an inner product between  $C$  and a random *rank-1* matrix  $\mathbf{r} \mathbf{r}^T$ . In the context of compressed sensing, this means that our method does not fit into the framework of Gaussian (or sub-Gaussian) random measurements [27], because of this rank-1 structure (or more concretely, because the matrix  $\mathbf{r} \mathbf{r}^T$  only involves  $n$  rather than  $n^2$  independent random variables).

In the context of sparse Hamiltonian estimation, this means that the theoretical analysis of our method requires different, considerably stronger techniques than those shown in [5, 6]. (In particular, our method requires probabilistic proof techniques that account for correlations between related random variables, such as “generic chaining” [28–30], in contrast to simpler techniques that neglect such correlations, such as the union bound used in [5, 6].)

Instead, our method turns out to have surprising connections to random Fourier measurements (and random measurements in incoherent bases) in compressed sensing [28, 29, 31, 32] (see also [27, 30, 33]). The key observation is that the off-diagonal elements of the matrix  $\mathbf{r} \mathbf{r}^T$  (subject to a suitable normalization condition) are random variables that are *centered* around 0 and *bounded* independent of the dimension  $n$ . These properties do not hold for the diagonal elements of  $\mathbf{r} \mathbf{r}^T$ , but those diagonal elements are irrelevant, because we are only trying to detect the off-diagonal part of  $C$ , which encodes correlations between *different* qubits. These observations imply that our method fits into the theoretical framework of compressed sensing using “bounded orthonormal systems” [28–32].

Using this theoretical framework, we prove several results about the accuracy of our compressed sensing method, and the physical resources needed to implement it in an experiment. In Section V, we show two different recovery guarantees for our method. First, we prove a “RIPless” recovery guarantee (Section V E), which shows that our method can reconstruct  $C$  accurately when the number of measurement settings is  $m = O(s \log n)$ , just slightly above the information-theoretic lower bound. (Here,

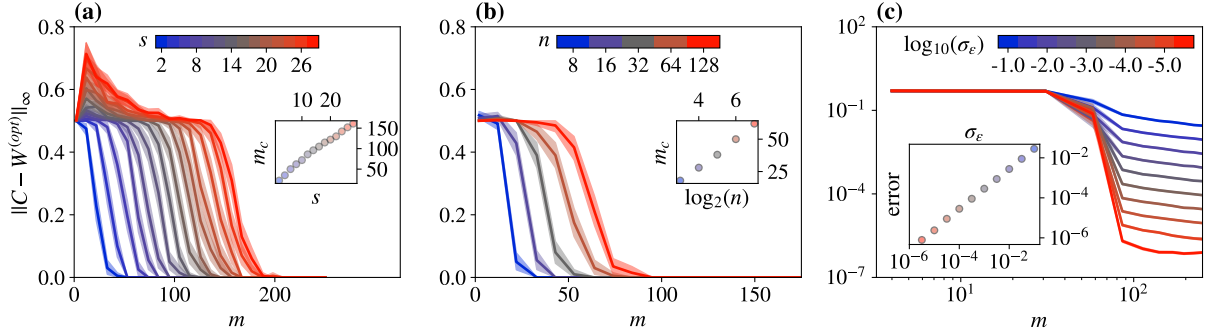


FIG. I.2. Scaling of the reconstruction error  $\|W^{(opt)} - C\|_\infty$  under various circumstances. Here  $W^{(opt)}$  denotes the estimate of  $C$  obtained via compressed sensing. The solid lines are the average errors over 100 random instances of the problem, and the shaded region is their 95% confidence interval obtained by bootstrapping. (a) The reconstruction error as a function of the number of measurement settings  $m$  (assuming noiseless measurements) for various values of sparsity  $s$  (and  $n = 64$  qubits). The errors go through a phase transition whose location  $m_c$  scales linearly with  $s$ . (b) The reconstruction error as a function of the number of measurement settings  $m$  (assuming noiseless measurements) for various numbers of qubits  $n$  (and sparsity  $s = 12$ ). The phase transition point  $m_c$  scales logarithmically with  $n$ . (c) The reconstruction error as a function of the number of measurement settings  $m$ , for different values of added noise strength  $\sigma_\epsilon$ , with fixed parameters  $(n, s) = (64, 12)$ . The inset shows that the recovery errors (when  $m > m_c$ , i.e., after the transition point) scale linearly with  $\sigma_\epsilon$ , as expected. See Section IV for details.

the ‘‘RIP’’ refers to the restricted isometry property, a standard proof technique in the theory of compressed sensing.)

Second, we prove a ‘‘RIP-based’’ recovery guarantee (Section VF), which shows that the recovery of  $C$  using our method is highly robust to noise in the measurements, provided that  $m$  is slightly larger, say  $m = O(s \log^4 n)$ . In addition, the ‘‘RIP-based’’ result shows *universal* recovery, meaning that a sin-

gle fixed measurement operator  $\Phi$  is capable of recovering all possible sparse matrices  $C$  (up to some unavoidable error due to the noise in the measurements).

In Section VI, we study the performance of our compressed sensing method, and compare it to the naive method where one measures each element of  $C$  independently. We can make a rigorous comparison, since we have error bounds for each of these methods. These can be summarized as follows:

$$\text{Naive: } \|\hat{C} - C\|_F \leq O(\sqrt{n}(\delta_1 + \delta_2)\|\text{diag}(C)\|_2 + \delta_2\|C'\|_F), \quad (\text{I.4})$$

$$\text{RIP-based: } \|W^{(opt)} - C\|_F \leq O(\sqrt{n}\delta_1\|\text{diag}(C)\|_2 + \delta_2(\sqrt{n}\|\text{diag}(C)\|_2 + \sqrt{2s}\|C'\|_F)), \quad (\text{I.5})$$

$$\text{RIPless: } \|W^{(opt)} - C\|_F \leq O(s \log^{5/2}(n)[\delta_1\|\text{diag}(C)\|_\infty \sqrt{sn \log(n)} + \delta_2\sqrt{n}\|\text{diag}(C)\|_2 + \delta_2\sqrt{2s}\|C'\|_F]). \quad (\text{I.6})$$

Here  $\hat{C}$  is the estimate of  $C$  using the naive method, and  $W^{(opt)}$  is the estimate of  $C$  using compressed sensing. (This is a simplification of the notation used in Section VI, which defines separate estimators for the diagonal and off-diagonal parts of  $C$ .) We write  $\text{diag}(C)$  to denote the diagonal of  $C$ , and  $C'$  to denote the off-diagonal part of  $C$ . We set

$m = O(s \log^4 n)$  for the RIP-based bound, and  $m = O(s \log n)$  for the RIPless bound. We then use (II.30) to bound the error in estimating  $\text{diag}(C)$ , and (II.31), (V.48) and (V.39) to bound the error in estimating  $C'$ . Finally,  $\delta_1$  and  $\delta_2$  are parameters that control the accuracy of the single-qubit and multi-qubit spectroscopy procedures. Using this theory,

Reconstruction method	# meas. settings $m$	Total sample complexity
Naive method	$O(n^2)$	$O(n^3/\delta^2)$
CS (RIP-based)	$O(s \log^4 n)$	$O(\max(n, s)^2 \log^4(n)/\delta^2)$
CS (RIPless*)	$O(s \log n)$	$O(s^3 \max(n, s) \log^6(n)/\delta^2)$

FIG. I.3. Sample complexity of different methods for reconstructing a correlation matrix  $C$ , of size  $n \times n$ , with  $2s$  nonzero elements off the diagonal. The naive method is to measure each element of  $C$  separately. “CS” refers to the compressed sensing method, and “RIP-based” and “RIPless” refer to different analytical bounds on the accuracy of the reconstruction of  $C$ . The asterisk (\*) indicates that the results using the RIPless bound hold under a technical assumption that the diagonal of  $C$  does not have any unusually large elements, see equation (VI.10). The different methods are parameterized in such a way that they reconstruct the diagonal of  $C$  up to an additive error of size  $\delta \|\text{diag}(C)\|_2$ , and they reconstruct the off-diagonal part of  $C$  up to an additive error of size  $\delta \|C\|_F$ . Each method makes use of single-qubit spectroscopy (with  $n$  experimental configurations or “measurement settings”), as well as multi-qubit spectroscopy (with  $m$  measurement settings, where  $m$  varies between  $O(n^2)$  and  $O(s \log n)$ ). The total sample complexity, shown in the table, includes both single-qubit and multi-qubit spectroscopy; for a more detailed accounting, see Figure VI.1. For the CS method, the number of measurement settings can be as low as  $m = O(s \log n)$ , using the RIPless bound, but the best sample complexity is achieved when  $m = O(s \log^4 n)$ , using the RIP-based bound. This sample complexity,  $O(\max(n, s)^2 \log^4(n))$ , compares favorably with the sample complexity of the naive method, which is  $O(n^3)$ ; we see that compressed sensing has an advantage over the naive method whenever  $s \leq O(n^{3/2}/\log^2 n)$ . See Section VI for details.

we can prove rigorous bounds on the sample complexity that is required for each method to reconstruct  $C$  with comparable accuracy. The results are summarized in Figure I.3.

For the compressed sensing method, the number of measurement settings can be as low as  $m = O(s \log n)$ , using the RIPless bound, but the best sample complexity is achieved when  $m = O(s \log^4 n)$ , using the RIP-based bound. This sample complexity,  $O(\max(n, s)^2 \log^4(n))$ , compares favorably with the sample complexity of the naive method, which is  $O(n^3)$ . Thus we see that compressed sensing has an advantage over the naive method whenever  $s \leq O(n^{3/2}/\log^2 n)$ .

Our method has some similarities with more recent work on phase retrieval, particularly the PhaseLift algorithm, although there are significant differences [34–38] (see also [8, 39]). We discuss this in detail in Section III B. We are not aware of any previous work on phase retrieval that directly addresses the situation studied in this paper; however, it is an interesting question whether techniques based on phase retrieval can be used to re-derive or improve on our results.

Our method has a second novel feature, which concerns the estimation of the decay rate  $\Gamma_{\mathbf{ab}} = 2\mathbf{r}^T C \mathbf{r}$ . In Sections II B and VII, we show a procedure for estimating  $\Gamma_{\mathbf{ab}}$  with high precision (i.e., with small multiplicative error), by allowing the system to evolve for a time  $t \sim 1/\Gamma_{\mathbf{ab}}$ . Here the time  $t$  is chosen by an adaptive procedure that starts with

an initial guess  $\tau_0$  and converges (provably) after  $\sim |\log(\Gamma_{\mathbf{ab}}\tau_0)|$  steps (see Fig. VII.1). To achieve high precision, we treat the time evolution operator  $e^{t\mathcal{L}}$  exactly, in contrast to previous work on sparse Hamiltonian estimation [5, 6], which used a linear approximation that is only valid for short times  $t$ .

This feature of our method is helpful in experimental setups where measurements are time-consuming and the values of  $\Gamma_{\mathbf{ab}}$  span several orders of magnitude, thus making it difficult to determine the appropriate evolution time and accurately measure the decay rates using conventional methods. This feature can also be used as a standalone technique to estimate the relaxation ( $T_1$ ) and decoherence ( $T_2$ ) times of any quantum system [25].

In Section VIII, we show that since our method relies on estimating exponential decay rates, it can be made at least partially robust to state-preparation-and-measurement (SPAM) errors (see Fig. VIII.1). This follows by some of the same approaches used in randomized benchmarking and gate set tomography [40–42].

Finally in Section IX we sketch a generalization of our method that includes unitary evolution according to some Hamiltonian  $H_S$ , and allows the matrix  $C$  to have complex entries  $c_{jk}$ . We show how a similar approach can be used to estimate the Hamiltonian  $H_S$ , as well as the imaginary part of  $c_{jk}$ .

It is interesting to compare our method with the recent work of [10] on estimation of sparse Pauli channels. At a technical level, these two works are

very different: they are measuring different types of noise (correlated dephasing versus Pauli errors), using different types of measurements (generalized Ramsey spectroscopy versus quantum Clifford circuits), and different reconstruction algorithms (convex optimization over a continuous domain, versus a combinatorial “peeling decoder”).

But from a broader perspective, these two methods do share certain general features. Both methods assume that the noise is sparse (albeit with very different mathematical representations), meaning that the noise is supported on a subset  $S$  of the domain, where  $S$  is small, but unknown. Both methods avoid making additional assumptions about the structure of  $S$  (e.g., in our work,  $S$  is allowed to contain long-range correlations, and in [10],  $S$  is allowed to contain high-weight Pauli errors). Finally, both methods utilize measurements of “decay rates” (albeit with very different types of experiments), in order to obtain results that are robust in the presence of state preparation and measurement (SPAM) errors.

This suggests that similar techniques, for SPAM-robust estimation of sparse noise models, can be used to characterize other kinds of correlated noise processes in many-body quantum systems.

### A. Notation

In this paper we use the following notation: Vectors are written in boldface, and matrices are denoted by capital letters. For a vector  $\mathbf{v}$ ,  $\|\mathbf{v}\|_p$  denotes the  $\ell_p$  norm (we will be mainly interested in the cases  $p = 1, 2, \infty$ ).

For a matrix  $M$ ,  $\|M\|_F = (\sum_{jk} |M_{jk}|^2)^{1/2}$  denotes the Frobenius norm (i.e., the Schatten 2-norm, or the  $\ell_2$  norm of the vector containing the entries of  $M$ ),  $\|M\|$  denotes the operator norm (i.e., the Schatten  $\infty$ -norm),  $\|M\|_{\text{tr}}$  denotes the trace norm (i.e., the Schatten 1-norm, or the nuclear norm), and  $\|M\|_{\ell_1} = \sum_{jk} |M_{jk}|$  denotes the  $\ell_1$  norm of the vector containing the entries of  $M$ .

Given an  $n \times n$  matrix  $M$ , let  $\text{uvec}(M) = (M_{ij})_{i < j}$  denote the vector (of dimension  $n(n-1)/2$ ) containing the entries in the upper-triangular part of  $M$ , excluding the diagonal.

Asymptotic bounds are written using big-O notation, such as  $O(s \log n)$ . Polylogarithmic factors are written in a compact way as follows:  $\log^c n = (\log n)^c$ . A “universal constant” is a quantity whose value is fixed once and for all, and does not depend on any other variable.

Statistical estimators are written with a hat super-

script, e.g.,  $\hat{\Gamma}$  is a random variable that represents an estimator for some unknown quantity  $\Gamma$ . For a random variable  $\hat{\Gamma}$ ,  $\|\hat{\Gamma}\|_{\psi_2}$  denotes the sub-Gaussian norm, and  $\|\hat{\Gamma}\|_{\psi_1}$  denotes the subexponential norm, in the sense of [27].

## II. GENERALIZED RAMSEY SPECTROSCOPY

We begin by describing a general form of Ramsey spectroscopy using entangled states on multiple qubits. We will also describe a simple method for directly measuring the correlation matrix  $C$ , by performing spectroscopy on every pair of qubits. This simple method serves as a baseline for measuring the performance of our compressed sensing method, which we will introduce in Section III.

Here we assume that the entries in the matrix  $C$  are real (i.e., with imaginary part equal to zero). This holds true in a number of important cases, for instance, when the qubits are coupled to a bath at high temperature (see Appendix A). When  $C$  is complex, it can be handled using a generalization of our method, described in Section IX.

Note that physical constraints imply that  $C$  is positive semidefinite [21]; hence we have  $C = C^T$  in the real case, and  $C = C^\dagger$  in the complex case.

### A. Dephasing of Entangled States

We begin by describing a procedure that allows us to measure certain linear functions of the correlation matrix  $C$ . This procedure is very general, and includes single- and two-qubit Ramsey spectroscopy as special cases. Consider an  $n$ -qubit state of the form

$$|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle) \in (\mathbb{C}^2)^{\otimes n}, \quad (\text{II.1})$$

where  $\mathbf{a}, \mathbf{b} \in \{0, 1\}^n$ ,  $\mathbf{a} \neq \mathbf{b}$ ,  $|\mathbf{a}\rangle = |a_1, a_2, \dots, a_n\rangle$  and  $|\mathbf{b}\rangle = |b_1, b_2, \dots, b_n\rangle$ . By choosing  $\mathbf{a}$  and  $\mathbf{b}$  appropriately, one can make  $|\psi_{\mathbf{ab}}\rangle$  be a single-qubit  $|+\rangle$  state, a two-qubit Bell state, or a many-qubit Greenberger-Horne-Zeilinger state (GHZ state) (while the other qubits are in a tensor product of  $|0\rangle$  and  $|1\rangle$  states).

Say we prepare the state  $|\psi_{\mathbf{ab}}\rangle$ , then allow it to evolve for time  $t$  under the Lindbladian (I.1). Let  $\rho(t)$  be the resulting density matrix. As can be seen in Eq. (I.2), the coherences (that is, the off-diagonal elements  $|\mathbf{a}\rangle\langle\mathbf{b}|$ ) of  $\rho(t)$  decay as  $\exp(-\Gamma_{\mathbf{ab}}t)$ , where

the decay rate  $\Gamma_{\mathbf{ab}} \in \mathbb{R}$  is defined so that  $\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = -\Gamma_{\mathbf{ab}}|\mathbf{a}\rangle\langle\mathbf{b}|$ . This decay rate can be estimated by allowing the system to evolve for a suitable amount of time  $t$ , and then measuring in the basis  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle \pm |\mathbf{b}\rangle)$  (see Section II B for details).

The decay rate  $\Gamma_{\mathbf{ab}}$  tells us a certain linear function of the correlation matrix  $C$ , which can be written explicitly as follows. Let  $\alpha_i = (-1)^{a_i}$  and  $\beta_i = (-1)^{b_i}$  denote the expectation values of  $Z_i$  corresponding to the states  $|\mathbf{a}\rangle$  and  $|\mathbf{b}\rangle$ , respectively. In addition, define the vectors  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$ . We can then see that

$$\Gamma_{\mathbf{ab}} = - \sum_{ij} c_{ij} [\alpha_i \beta_j - \frac{1}{2} \alpha_i \alpha_j - \frac{1}{2} \beta_i \beta_j] \quad (\text{II.2})$$

$$= 2\mathbf{r}^T C \mathbf{r}, \quad (\text{II.3})$$

where we recall the definition of  $\mathbf{r}$ ,

$$\mathbf{r} = \mathbf{b} - \mathbf{a}, \quad (\text{II.4})$$

we note that  $\mathbf{r} = \frac{\boldsymbol{\alpha} - \boldsymbol{\beta}}{2}$ , and we use the fact that  $C = C^T$  to symmetrize the equation.

Single-qubit Ramsey spectroscopy (Fig. I.1(c)) is a special case of this procedure, where we set  $\mathbf{a} = (0, 0, \dots, 0)$  and  $\mathbf{b} = (0, \dots, 0, 1, 0, \dots, 0)$  (where the 1 appears in the  $j$ 'th position). Then  $|\psi_{\mathbf{ab}}\rangle$  is a  $|+\rangle$  state on the  $j$ 'th qubit, and  $\Gamma_{\mathbf{ab}} = 2c_{jj}$  tells us the rate of dephasing on the  $j$ 'th qubit.

Two-qubit generalized spectroscopy is another special case, where we set  $\mathbf{a} = (0, 0, \dots, 0)$  and  $\mathbf{b} = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$  (where the 1's appear in the  $j$ 'th and  $k$ 'th positions). Then  $|\psi_{\mathbf{ab}}\rangle$  is a maximally-entangled state on qubits  $j$  and  $k$ , and  $\Gamma_{\mathbf{ab}} = 2(c_{jj} + c_{jk} + c_{kj} + c_{kk})$  gives us information about the rate of correlated dephasing on qubits  $j$  and  $k$ .

## B. Estimating Decay Rates

There are many possible ways to estimate the decay rate  $\Gamma_{\mathbf{ab}}$ . For concreteness, we describe one simple and rigorous method here:

1. Choose some evolution time  $t \geq 0$  such that  $\frac{1}{2} \leq \Gamma_{\mathbf{ab}} t \leq 2$ .

This can be done in various ways, for instance, by starting with an initial guess  $t = \tau_0$  and performing binary search, using  $\sim |\log(\Gamma_{\mathbf{ab}} \tau_0)|$  trials of the experiment (see Section VII for details).

2. Repeat the following experiment  $N_{\text{trials}}$  times: (we can set  $N_{\text{trials}}$  using equation (II.11) below)

- (a) Prepare the state  $|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ , allow the state to evolve for time  $t$ , then measure in the basis  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle \pm |\mathbf{b}\rangle)$ .

Let  $N_+$  and  $N_-$  be the number of  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$  and  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle - |\mathbf{b}\rangle)$  outcomes, respectively. Note that the probabilities of these outcomes are given by  $P_+ = \frac{1}{2}(1 + e^{-\Gamma_{\mathbf{ab}} t})$  and  $P_- = \frac{1}{2}(1 - e^{-\Gamma_{\mathbf{ab}} t})$ .

3. Define:

$$\Delta = \frac{N_+ - N_-}{N_{\text{trials}}}. \quad (\text{II.5})$$

Note that  $\Delta$  is an unbiased estimator for  $P_+ - P_-$ , that is,  $\mathbb{E}(\Delta) = P_+ - P_- = e^{-\Gamma_{\mathbf{ab}} t}$ . This motivates our definition of an estimator for  $\Gamma_{\mathbf{ab}}$ :

$$\hat{\Gamma}_{\mathbf{ab}} = -\frac{1}{t} \ln(\Delta). \quad (\text{II.6})$$

We now state some bounds on the accuracy of the estimator  $\hat{\Gamma}_{\mathbf{ab}}$ . To do this, we introduce the notion of a sub-Gaussian random variable (roughly speaking, a random variable whose moments and tail probabilities behave like those of a Gaussian distribution) [27]. Formally, we say that a real-valued random variable  $X$  is *sub-Gaussian* if there exists a real number  $K_2$  such that,

$$\text{for all } p \geq 1, (\mathbb{E}(|X|^p))^{1/p} \leq K_2 \sqrt{p}. \quad (\text{II.7})$$

The *sub-Gaussian norm* of  $X$ , denoted  $\|X\|_{\psi_2}$ , is defined to be the smallest choice of  $K_2$  in (II.7), i.e.,

$$\|X\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathbb{E}(|X|^p))^{1/p}. \quad (\text{II.8})$$

In addition, it is known that  $X$  is sub-Gaussian if and only if there exists a real number  $K_1$  such that,

$$\text{for all } t \geq 0, \Pr[|X| > t] \leq \exp(1 - t^2/K_1^2). \quad (\text{II.9})$$

The smallest choice of  $K_1$  in (II.9) is equivalent to the sub-Gaussian norm  $\|X\|_{\psi_2}$ , in the following sense: there is a universal constant  $c$  such that, for all sub-Gaussian random variables  $X$ , the smallest choice of  $K_1$  in (II.9) differs from  $\|X\|_{\psi_2}$  by at most a factor of  $c$ .

We now show that  $\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}$  is a sub-Gaussian random variable, whose sub-Gaussian norm is bounded by

$$\|\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}\|_{\psi_2} \leq C_0 \frac{\Gamma_{\mathbf{ab}}}{\sqrt{N_{\text{trials}}}}, \quad (\text{II.10})$$

where  $C_0$  is a universal constant. (This  $1/\sqrt{N_{\text{trials}}}$  scaling is familiar from classical statistics, and is not novel. The novelty of this paper will appear later, when we analyze the sample complexity of our compressed sensing estimators, in Section VI.)

In particular, the accuracy of  $\hat{\Gamma}_{\mathbf{ab}}$  can be controlled by setting  $N_{\text{trials}}$  appropriately: for any  $\delta > 0$  and  $\epsilon > 0$ , if we set

$$N_{\text{trials}} \geq \frac{2}{\delta^2} \ln\left(\frac{2}{\epsilon}\right), \quad (\text{II.11})$$

then  $\hat{\Gamma}_{\mathbf{ab}}$  satisfies the following error bound: with probability at least  $1 - \epsilon$ ,

$$|\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}| \leq 2\delta e^2 \Gamma_{\mathbf{ab}}. \quad (\text{II.12})$$

This shows that the error in  $\hat{\Gamma}_{\mathbf{ab}}$  is at most a small fraction of the true value of  $\Gamma_{\mathbf{ab}}$ , independent of the magnitude of  $\Gamma_{\mathbf{ab}}$ .

We can also use this to derive an error bound that involves  $\hat{\Gamma}_{\mathbf{ab}}$  rather than  $\Gamma_{\mathbf{ab}}$ , and hence can be computed from the observed value of  $\hat{\Gamma}_{\mathbf{ab}}$ . To show such an error bound, use the triangle inequality to write  $|\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}| \leq 2\delta e^2 (|\hat{\Gamma}_{\mathbf{ab}}| + |\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}|)$ , and divide by  $(1 - 2\delta e^2)$  to get:

$$|\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}| \leq \frac{2\delta e^2}{1 - 2\delta e^2} |\hat{\Gamma}_{\mathbf{ab}}|. \quad (\text{II.13})$$

It remains to prove (II.10) and (II.12). First use Hoeffding's inequality to show that  $\Delta$  is close to its expectation value:

$$\Pr[|\Delta - (P_+ - P_-)| \geq \delta] \leq 2 \exp(-N_{\text{trials}} \delta^2 / 2). \quad (\text{II.14})$$

When  $|\Delta - (P_+ - P_-)| \leq \delta$ , we can bound the error in  $\hat{\Gamma}_{\mathbf{ab}}$  as follows: (using the fact that  $P_+ - P_- = e^{-\Gamma_{\mathbf{ab}} t} \geq e^{-2}$  and  $t \geq \frac{1}{2\Gamma_{\mathbf{ab}}}$ )

$$\begin{aligned} -\hat{\Gamma}_{\mathbf{ab}} &\leq \frac{1}{t} \ln(P_+ - P_- + \delta) \\ &\leq \frac{1}{t} [\ln(P_+ - P_-) + \delta e^2] \\ &\leq -\Gamma_{\mathbf{ab}} + 2\Gamma_{\mathbf{ab}} \delta e^2, \end{aligned} \quad (\text{II.15})$$

$$\begin{aligned} -\hat{\Gamma}_{\mathbf{ab}} &\geq \frac{1}{t} \ln(P_+ - P_- - \delta) \\ &\geq \frac{1}{t} [\ln(P_+ - P_-) - \delta e^2] \\ &\geq -\Gamma_{\mathbf{ab}} - 2\Gamma_{\mathbf{ab}} \delta e^2. \end{aligned} \quad (\text{II.16})$$

Hence we have:

$$\Pr[|\hat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}| \geq 2\delta e^2 \Gamma_{\mathbf{ab}}] \leq 2 \exp(-N_{\text{trials}} \delta^2 / 2). \quad (\text{II.17})$$

This implies (II.10) and (II.12).

### C. Direct Estimation of the Correlation Matrix

There is a simple way to estimate the correlation matrix  $C$  directly, by performing single-qubit spectroscopy to measure the diagonal elements  $c_{jj}$ , and performing two-qubit spectroscopy to measure the off-diagonal elements  $c_{jk}$ . We describe this method here. We will use this method as a baseline, to measure the performance of the compressed sensing method that we will introduce in Section III.

For simplicity, we consider the case where  $C$  is real. Since  $C$  is positive definite, this implies that  $c_{jj} \geq 0$  and  $c_{jk} = c_{kj}$ .

First, we estimate the diagonal elements  $c_{jj}$ , for  $j = 1, \dots, n$ , as follows:

1. Let  $\mathbf{a} = (0, 0, \dots, 0)$  and  $\mathbf{b} = (0, \dots, 0, 1, 0, \dots, 0)$  (where the 1 appears in the  $j$ 'th position).
2. Construct an estimate  $\hat{\Gamma}_{\mathbf{ab}}$  of the decay rate  $\Gamma_{\mathbf{ab}} = 2c_{jj}$  (for instance, by using the procedure in Section IIB). Define  $g_j = \hat{\Gamma}_{\mathbf{ab}}/2$ .

To write this in a compact form, we define

$$\text{diag}(C) = (c_{11}, \dots, c_{nn}), \quad (\text{II.18})$$

that is, the diagonal of  $C$ . We then define

$$\mathbf{g} = (g_1, \dots, g_n), \quad (\text{II.19})$$

and we view this as an estimator for  $\text{diag}(C)$ .

Next, we estimate the off-diagonal elements  $c_{jk}$ , for  $1 \leq j < k \leq n$ , as follows:

1. Let  $\mathbf{a} = (0, 0, \dots, 0)$  and  $\mathbf{b} = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$  (where the 1's appears in the  $j$ 'th and  $k$ 'th positions).
2. Construct an estimate  $\hat{\Gamma}_{\mathbf{ab}}$  of the decay rate  $\Gamma_{\mathbf{ab}} = 2(c_{jj} + 2c_{jk} + c_{kk})$  (for instance, by using the procedure in Section IIB). Define  $h_{jk} = \frac{1}{4} \hat{\Gamma}_{\mathbf{ab}} - \frac{1}{2} g_j - \frac{1}{2} g_k$ .

To write this in a compact form, we define  $C'$  to be the matrix  $C$  with the diagonal entries replaced by zeroes,

$$C'_{jk} = \begin{cases} c_{jk} & \text{if } j \neq k, \\ 0 & \text{if } j = k. \end{cases} \quad (\text{II.20})$$



We call this the ‘‘off-diagonal part’’ of  $C$ . We then construct an estimator  $\widehat{C}'$  for  $C'$ , as follows:

$$\widehat{C}'_{jk} = \begin{cases} h_{jk} & \text{if } j < k, \\ h_{kj} & \text{if } j > k, \\ 0 & \text{if } j = k. \end{cases} \quad (\text{II.21})$$

We can analyze the accuracy of these estimators as follows. Choose two parameters  $\delta_1$  and  $\delta_2$ . Suppose that, during the measurement of the diagonal elements  $c_{jj}$ , the decay rates  $\Gamma_{\mathbf{ab}}$  are estimated with accuracy

$$\|\widehat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}\|_{\psi_2} \leq \delta_1 \Gamma_{\mathbf{ab}}, \quad (\text{II.22})$$

and during the measurement of the off-diagonal elements  $c_{jk}$ , the decay rates  $\Gamma_{\mathbf{ab}}$  are estimated with accuracy

$$\|\widehat{\Gamma}_{\mathbf{ab}} - \Gamma_{\mathbf{ab}}\|_{\psi_2} \leq \delta_2 \Gamma_{\mathbf{ab}}. \quad (\text{II.23})$$

Bounds of the form (II.22) and (II.23) can be obtained from equation (II.10), by setting  $N_{\text{trials}} \sim 1/\delta_1^2$  and  $N_{\text{trials}} \sim 1/\delta_2^2$ , respectively. Here, we are neglecting to count those trials of the experiment that are used to choose the evolution time  $t$ , because the number of those trials grows only logarithmically with  $\Gamma_{\mathbf{ab}}$ . We allow  $\delta_1$  and  $\delta_2$  to be different, because in many experimental scenarios, measurements of  $c_{jj}$  take less time than measurements of  $c_{jk}$ .

Using (II.22) and (II.23), we can easily show bounds on the accuracy of  $g_j$  and  $h_{jk}$ :

$$\|g_j - c_{jj}\|_{\psi_2} \leq \delta_1 c_{jj}, \quad (\text{II.24})$$

$$\|h_{jk} - c_{jk}\|_{\psi_2} \leq \delta_2 c_{jk} + \frac{1}{2}(\delta_1 + \delta_2)(c_{jj} + c_{kk}). \quad (\text{II.25})$$

These bounds on the sub-Gaussian norm imply bounds on the moments, such as  $\mathbb{E}|X| \leq \|X\|_{\psi_2}$  and  $\mathbb{E}(X^2) \leq 2\|X\|_{\psi_2}^2$  (see [27] for details).

We now use these results to bound the accuracy of the estimators  $\mathbf{g}$  and  $\widehat{C}'$ . For  $\mathbf{g}$ , we have the following bounds, using the  $\ell_1$  and  $\ell_2$  norms:

$$\mathbb{E}(\|\mathbf{g} - \text{diag}(C)\|_1) \leq \delta_1 \|\text{diag}(C)\|_1, \quad (\text{II.26})$$

$$\mathbb{E}(\|\mathbf{g} - \text{diag}(C)\|_2^2) \leq 2\delta_1^2 \|\text{diag}(C)\|_2^2. \quad (\text{II.27})$$

For  $\widehat{C}'$ , we have the following bounds, using the  $\ell_1$  vector norm and the Frobenius matrix norm:

$$\begin{aligned} & \mathbb{E}(\|\widehat{C}' - C'\|_{\ell_1}) \\ & \leq 2 \sum_{j < k} (\delta_2 c_{jk} + \frac{1}{2}(\delta_1 + \delta_2)(c_{jj} + c_{kk})) \\ & = \delta_2 \sum_{j \neq k} c_{jk} + \frac{1}{2}(\delta_1 + \delta_2) \sum_{j \neq k} (c_{jj} + c_{kk}) \\ & \leq (n-1)(\delta_1 + \delta_2) \|\text{diag}(C)\|_{\ell_1} + \delta_2 \|C'\|_{\ell_1}, \end{aligned} \quad (\text{II.28})$$

$$\begin{aligned} & \mathbb{E}(\|\widehat{C}' - C'\|_F^2) \\ & \leq 2 \sum_{j < k} 2[\delta_2 c_{jk} + \frac{1}{2}(\delta_1 + \delta_2)(c_{jj} + c_{kk})]^2 \\ & \leq 2 \sum_{j < k} 6[\delta_2^2 c_{jk}^2 + \frac{1}{4}(\delta_1 + \delta_2)^2 c_{jj}^2 + \frac{1}{4}(\delta_1 + \delta_2)^2 c_{kk}^2] \\ & \leq 6\delta_2^2 \|C'\|_F^2 + \frac{3}{2}(\delta_1 + \delta_2)^2 \sum_{j \neq k} (c_{jj}^2 + c_{kk}^2) \\ & \leq 3(n-1)(\delta_1 + \delta_2)^2 \|\text{diag}(C)\|_2^2 + 6\delta_2^2 \|C'\|_F^2. \end{aligned} \quad (\text{II.29})$$

Note that, in the second step of (II.29), we used the fact that for any real numbers  $a_1, a_2, a_3$ ,  $(a_1 + a_2 + a_3)^2 \leq 3(a_1^2 + a_2^2 + a_3^2)$ .

These are bounds on the expected error of  $\mathbf{g}$  and  $\widehat{C}'$ . One can then use Markov’s inequality to prove bounds that hold with high probability. For instance, using (II.27), we get that with probability at least  $1 - \eta$ ,

$$\|\mathbf{g} - \text{diag}(C)\|_2 \leq \frac{1}{\sqrt{\eta}} \sqrt{2\delta_1} \|\text{diag}(C)\|_2, \quad (\text{II.30})$$

and using (II.29), we get that with probability at least  $1 - \eta$ ,

$$\begin{aligned} & \|\widehat{C}' - C'\|_F \\ & \leq \frac{1}{\sqrt{\eta}} [3(n-1)(\delta_1 + \delta_2)^2 \|\text{diag}(C)\|_2^2 + 6\delta_2^2 \|C'\|_F^2]^{1/2} \\ & \leq \frac{1}{\sqrt{\eta}} [\sqrt{3}\sqrt{n-1}(\delta_1 + \delta_2) \|\text{diag}(C)\|_2 + \sqrt{6}\delta_2 \|C'\|_F]. \end{aligned} \quad (\text{II.31})$$

In fact, one can prove tighter bounds on the failure probability, by replacing Markov’s inequality with a sharp concentration bound for sub-Gaussian random variables [27]. However, these tighter bounds are not needed for the purposes of this paper.

The bounds (II.30) and (II.31) give a rough sense of how well this estimator performs. In particular, these bounds show that the error in estimating  $C'$  depends on the magnitude of  $\text{diag}(C)$ , as well as the magnitude of  $C'$ . This is due to the fact that our procedure for estimating the off-diagonal matrix element  $c_{jk}$  also involves the diagonal matrix elements  $c_{jj}$  and  $c_{kk}$ . Later in the paper, we will use these bounds as a baseline to understand the performance of our compressed sensing estimators (see Section VI).

### III. LEARNING SPARSE CORRELATIONS VIA COMPRESSED SENSING

Our main contribution in this paper is an efficient method for learning the off-diagonal part of

the correlation matrix  $C$ , under the assumption that it is *sparse*, i.e., the part of  $C$  that lies above the diagonal has at most  $s$  nonzero elements, where  $s \ll n(n-1)/2$ . (Since  $C$  is Hermitian, it is sufficient to learn the part that lies above the diagonal; this then determines the part that lies below the diagonal.)

For simplicity, we first consider the special case where the entries in the matrix  $C$  are real (i.e., with zero imaginary part), which occurs in a number of physical situations (for example, when the system is coupled to a bath at high temperature, see Appendix A). Later in Section IX we will show how our method can be extended to handle complex matrices  $C$ .

Our method consists of two steps: first, we perform single-qubit Ramsey spectroscopy in order to learn the diagonal elements of  $C$ ; second, we apply techniques from compressed sensing (e.g., random measurements, and  $\ell_1$ -minimization) in order to recover the off-diagonal elements of  $C$ .

### A. Random Measurements of the Correlation Matrix

We now describe our method in more detail. First, we estimate each of the diagonal elements  $c_{jj}$ , for  $j = 1, \dots, n$ , using single-qubit Ramsey spectroscopy, as described in Fig. I.1(c). Let  $\mathbf{g} = (g_1, \dots, g_n) \in \mathbb{R}^n$  be the output of this procedure (this is the same notation used in Section II C). We can view  $\mathbf{g}$  as an estimate of a “sensing operator” that returns the diagonal elements of the matrix  $C$ ,

$$\mathbf{g} \approx \text{diag}(C) = (c_{11}, c_{22}, \dots, c_{nn}). \quad (\text{III.1})$$

(Note that  $c_{jj} \geq 0$ , since  $C$  is positive semidefinite.)

In order to estimate the off-diagonal part of  $C$ , we will use a compressed sensing technique, which involves a certain type of generalized Ramsey measurement with random GHZ-type states, see Fig. I.1(d). First, we choose a parameter  $m$ , which can be roughly  $m \sim s \log n$  or  $m \sim s \log^4 n$ , which controls the number of different measurements. (The particular choice of  $m$  is motivated by the theoretical recovery guarantees in Section V.) Now, for  $j = 1, \dots, m$ , we perform the following procedure:

1. Choose vectors  $\mathbf{a}, \mathbf{b} \in \{0, 1\}^n$  uniformly at random. As in equation (II.4), define

$$\mathbf{r} = \mathbf{b} - \mathbf{a}. \quad (\text{III.2})$$

2. Prepare the state  $|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ . This is a GHZ state on a subset of the qubits, with

some bit flips. It can be created by preparing those qubits  $i$  where  $a_i = b_i$  in the state  $|a_i\rangle$ , preparing those qubits  $i$  where  $a_i \neq b_i$  in a GHZ state, and applying a Pauli  $X$  operator on those qubits  $i$  where  $a_i > b_i$ . (This requires a quantum circuit of depth  $\lceil \log_2(n) \rceil + 2$ .)

3. Construct an estimate  $\hat{\Gamma}_{\mathbf{ab}}$  of the decay rate  $\Gamma_{\mathbf{ab}} = 2\mathbf{r}^T C \mathbf{r}$  (for instance, using the procedure in Section II B). Define  $h_j = \hat{\Gamma}_{\mathbf{ab}}$ .

Let  $\mathbf{h} = (h_1, \dots, h_m) \in \mathbb{R}^m$  be the output of the above procedure. Again, we can view  $\mathbf{h}$  as an estimate of a “sensing operator”  $\Phi : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^m$  that is applied to the matrix  $C$ ,

$$\mathbf{h} \approx \Phi(C) = [\Phi_j(C)]_{j=1, \dots, m} \quad (\text{III.3})$$

$$\Phi_j(C) = 2(\mathbf{r}^{(j)})^T C \mathbf{r}^{(j)} \quad (\text{III.4})$$

where  $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \dots, \mathbf{r}^{(m)} \in \{1, 0, -1\}^n$  are independent random vectors chosen from the same distribution as  $\mathbf{r}$  (described above). Note that  $\Phi_j(C) \geq 0$ , since  $C$  is positive semidefinite. The factor of 2 is chosen to ensure that  $\Phi$  has a certain isotropy property, which will be discussed in Section V D.

### B. Reconstructing the Correlation Matrix

We now show how to reconstruct the correlation matrix  $C \in \mathbb{R}^{n \times n}$ . We are promised that  $C$  is positive semidefinite, due to physical constraints [21], and its off-diagonal part is sparse, with at most  $s \ll n(n-1)/2$  nonzero elements above the diagonal. In general, this sparsity constraint leads to an optimization problem that is computationally intractable. However, in this particular case, this problem can be solved using a strategy from compressed sensing: given  $\mathbf{g} \approx \text{diag}(C) \in \mathbb{R}^n$  and  $\mathbf{h} \approx \Phi(C) \in \mathbb{R}^m$ , we will recover  $C$  by solving a convex optimization problem, where we minimize the  $\ell_1$  (vector) norm of the off-diagonal part of the matrix. We will show that this strategy succeeds when  $m \geq c_0 s \log n$ , and is highly robust to noise when  $m \geq c_0 s \log^4 n$ , where  $c_0$  is some universal constant.

We now describe this approach in more detail. First, we consider the case where the measurements are noiseless, i.e.,  $\mathbf{g} = \text{diag}(C)$  and  $\mathbf{h} = \Phi(C)$ . We solve the following convex optimization problem:

$$\text{Find } W \in \mathbb{R}^{n \times n} \text{ that minimizes } \sum_{i \neq j} |W_{ij}|, \quad (\text{III.5})$$

such that:

$$\text{diag}(W) = \mathbf{g}, \quad (\text{III.6})$$

$$\Phi(W) = \mathbf{h}, \quad (\text{III.7})$$

$$W \succeq 0. \quad (\text{III.8})$$

Here,  $W \succeq 0$  means that  $W$  is positive semidefinite, which implies that  $W = W^T$ . As a sanity check, note that  $W = C$  is a feasible solution to this problem. (Recall that  $C$  is positive semidefinite.)

We remark that this scheme bears some resemblance to the PhaseLift algorithm for phase retrieval [34–38]. In phase retrieval, one wishes to estimate an unknown vector  $\mathbf{x}$  from measurements of the form  $|\mathbf{r}^T \mathbf{x}|^2$ . The PhaseLift algorithm works by “lifting” the unknown vector  $\mathbf{x}$  to a matrix  $X = \mathbf{x}\mathbf{x}^T$ , so that the problem becomes one of learning a rank-1 matrix  $X$  from measurements of the form  $\mathbf{r}^T X \mathbf{r}$ ; then one solves a convex relaxation of this problem. In cases where the unknown vector  $\mathbf{x}$  is sparse (“compressive phase retrieval”), variants of the PhaseLift algorithm (as well as other approaches) can also be used [35–37].

The main difference between our method and PhaseLift is that, in our method, the unknown matrix  $C$  is almost always full-rank (because every qubit has a nonzero dephasing rate), whereas in PhaseLift, the unknown matrix  $X$  has rank 1. In our situation, physical constraints imply that  $C$  is positive semidefinite, so it can be factored as  $C = BB^T$ , which is superficially similar to  $X = \mathbf{x}\mathbf{x}^T$ ; however, an important difference is that  $B$  is a square matrix, whereas  $\mathbf{x}$  is a vector. Methods like PhaseLift have been extended to handle low-rank matrices  $X$ , albeit without taking advantage of sparsity [38], and it is an interesting question whether one can use this approach to re-derive or improve on our method, where sparsity plays a crucial role.

### C. Reconstruction from Noisy Measurements

In the case where the measurements of  $\mathbf{g}$  and  $\mathbf{h}$  are noisy, we need to modify the above convex optimization problem, by relaxing the constraints (III.6) and (III.7). This leads to some technical complications, due to the fact that we are reconstructing two variables that have different characteristics: the diagonal part of  $C$  (which is not sparse), and the off-diagonal part of  $C$  (which is sparse).

To deal with these issues, we propose two different ways of performing this reconstruction, when

the measurements are noisy: (1) simultaneous reconstruction of both parts of  $C$ , and (2) sequential reconstruction of the diagonal part of  $C$ , followed by the off-diagonal part of  $C$ . The former approach is arguably more natural, but the latter approach allows for more rigorous analysis of the accuracy of the reconstruction (see Section V).

Suppose we have bounds on the  $\ell_2$  norms of the noise terms (which we denote  $\mathbf{u}$  and  $\mathbf{v}$ ), that is,

$$\mathbf{g} = \text{diag}(C) + \mathbf{u}, \quad \|\mathbf{u}\|_2 \leq \epsilon_1, \quad (\text{III.9})$$

$$\mathbf{h} = \Phi(C) + \mathbf{v}, \quad \|\mathbf{v}\|_2 \leq \epsilon_2. \quad (\text{III.10})$$

(We do not assume anything about the distribution of  $\mathbf{u}$  and  $\mathbf{v}$ . We will describe how to set  $\epsilon_1$  and  $\epsilon_2$  below, for some typical measurement procedures.)

*Simultaneous reconstruction of the diagonal and off-diagonal parts of  $C$ :* Here we relax the constraints (III.6) and (III.7) in the simplest possible way, by replacing them with:

$$\|\text{diag}(W) - \mathbf{g}\|_2 \leq \epsilon_1, \quad (\text{III.11})$$

$$\|\Phi(W) - \mathbf{h}\|_2 \leq \epsilon_2. \quad (\text{III.12})$$

This leads to a convex optimization problem that attempts to reconstruct both the diagonal part of  $C$ , which is not necessarily sparse, and the off-diagonal part of  $C$ , which is assumed to be sparse. (Note that  $W = C$  is a feasible solution to this problem.) This method often works quite well in practice.

Unfortunately, the behavior of this reconstruction algorithm can be complicated, because it involves two different estimators (an  $\ell_1$ -regularized estimator for the off-diagonal part of  $C$ , and a least-squares estimator for the diagonal part of  $C$ ). These two estimators are coupled together (through the constraint on  $\Phi(W)$ , and the positivity constraint  $W \succeq 0$ ).

Therefore, this method can behave quite differently, depending on whether the dominant source of error is  $\mathbf{g}$  or  $\mathbf{h}$ . When  $\mathbf{g}$  is the dominant source of error, this method will behave like a least-squares estimator, whose accuracy scales according to the dimension  $n$ ; when  $\mathbf{h}$  is the dominant source of error, this method will behave like an  $\ell_1$ -regularized estimator, whose accuracy scales according to the sparsity  $s$  (neglecting log factors). From a theoretical point of view, this makes it more difficult to prove recovery guarantees for this method.

*Sequential reconstruction of the diagonal part of  $C$ , followed by the off-diagonal part of  $C$ :* In practice, one is often interested in the regime where  $\mathbf{g}$  is

known with high precision, and  $\mathbf{h}$  is the dominant source of error. This is because measurements of  $\mathbf{g}$  are relatively easy to perform, because they only require single-qubit state preparations and measurements; whereas measurements of  $\mathbf{h}$  are more costly, because they require the preparation and measurement of entangled states on many qubits. So measurements of  $\mathbf{g}$  can often be performed more quickly, and measurements on different qubits can be performed simultaneously in parallel; hence one can repeat the measurements more times, to obtain more accurate estimates of  $\mathbf{g}$ .

In this regime, it is natural to try to recover the diagonal part of  $C$  *directly* from  $\mathbf{g}$ , and then use  $\ell_1$ -minimization to recover *only* the off-diagonal part of  $C$ . This leads to a convex optimization problem which is arguably less natural, but it makes it easier to prove rigorous guarantees on the accuracy of the reconstruction of  $C$  (see Section V).

We now describe this approach in detail. We take the convex optimization problem ((III.5)-(III.8)) for the noiseless case, and we relax the last two constraints to get:

$$\text{Find } W \in \mathbb{R}^{n \times n} \text{ that minimizes } \sum_{i \neq j} |W_{ij}|, \quad (\text{III.13})$$

such that:

$$\text{diag}(W) = \mathbf{g}, \quad (\text{III.14})$$

$$\|\Phi(W) - \mathbf{h}\|_2 \leq \epsilon_2 + \epsilon_1 \sqrt{mn}, \quad (\text{III.15})$$

$$d(W, K_+) \leq \epsilon_1, \text{ and } W = W^T. \quad (\text{III.16})$$

Here  $K_+ = \{W' \in \mathbb{R}^{n \times n} \mid W' \succeq 0\}$  denotes the (real) positive semidefinite cone, and we define

$$d(W, K_+) = \min_{W' \in K_+} \|W - W'\|_F \quad (\text{III.17})$$

to be the minimum distance from  $W$  to a point  $W'$  in  $K_+$ , measured in Frobenius norm; note that this is a convex function. While this convex optimization problem looks complicated, it follows from a simple underlying idea: since the diagonal elements of  $W$  are fixed by the constraint (III.14), this is simply an  $\ell_1$ -regularized estimator for the sparse, off-diagonal part of  $W$ .

The attentive reader will notice two potential concerns with this approach. First, in general,  $C$  will not be a feasible solution to this convex optimization problem, since the diagonal elements of  $C$  will not satisfy (III.14). However, we claim that  $C$  lies close

to a feasible solution. To see this, let  $\tilde{C}$  be the matrix whose off-diagonal elements agree with  $C$ , and whose diagonal elements agree with  $\mathbf{g}$ . Then  $C$  is within distance  $\epsilon_1$  of  $\tilde{C}$  (in Frobenius norm), and we claim that  $\tilde{C}$  is a feasible solution. To see this, we can check that  $\tilde{C}$  satisfies the constraints (III.14), (III.15) and (III.16), since we have:

$$\begin{aligned} \|\Phi(\tilde{C}) - \Phi(C)\|_2 &= \|\Phi(\text{diag}(\mathbf{g} - \text{diag}(C)))\|_2 \\ &\leq (m \|\mathbf{g} - \text{diag}(C)\|_1^2)^{1/2} \\ &\leq \epsilon_1 \sqrt{mn}. \end{aligned} \quad (\text{III.18})$$

(Here, we wrote  $\tilde{C}$  in a compact form,  $\tilde{C} = C + \text{diag}(\mathbf{g} - \text{diag}(C))$ , where the  $\text{diag}(\cdot)$  notation has the following meaning: for a matrix  $M$ ,  $\text{diag}(M)$  is the vector containing the entries  $M_{jj}$  that lie along the diagonal of  $M$ ; and for a vector  $\mathbf{v} = (v_1, \dots, v_n)$ ,  $\text{diag}(\mathbf{v})$  is the diagonal matrix with  $v_1, \dots, v_n$  along the diagonal.)

Second, the reader will notice that the optimal solution  $W$  may violate the positivity constraint (III.8), making it un-physical. (Similar issues can arise when performing quantum state and process tomography.) However,  $W$  can be easily corrected to get a physically-admissible solution. This follows because equation (III.16) shows that  $W$  lies within distance  $\epsilon_1$  of a physically-admissible solution  $W' \succeq 0$ , and this solution  $W'$  can be obtained by truncating the negative eigenvalues of  $W$ .

Finally, we remark that there are different ways of relaxing the positivity constraint (III.8), and (III.16) is not the strongest possible choice. For instance, we could have used a stronger constraint than (III.16), such as:  $d'(W, K_+) \leq \epsilon_1$ , where we define  $d'(W, K_+) = \min_{W' \in K_+} \|\text{diag}(W - W')\|_2$ . However, the constraint (III.16) may be simpler to implement using numerical linear algebra software.

Since these are convex optimization problems, they can be solved efficiently (both in theory and in practice), for instance by using interior point algorithms. Nonetheless, some care is needed to ensure that these algorithms can scale up to solve very large instances of these problems. In particular, enforcing the positivity constraint (III.8), and its relaxed version (III.16), can be computationally expensive.

#### D. Omitting the Positivity Constraint

The theoretical analysis in Section V shows that  $C$  can be reconstructed by solving the convex optimization problem (III.13)-(III.16). We remark that

this analysis holds even *without* the positivity constraint (III.16). It is easy to check that the positivity constraint is not used in Section V, and indeed, most of the theory of compressed sensing applies to all sparse signals, not just positive ones, although positivity can be helpful in certain situations [43].

This observation has a practical consequence: by omitting the positivity constraint (III.16), one can make the convex optimization problem simpler, and thus easier to solve in practice (e.g., by using second-order cone programming, rather than semidefinite programming) [44]. One can then take the resulting solution, and project it onto the positive semidefinite cone, as is sometimes done in quantum state tomography [45, 46], without increasing the error (in Frobenius norm). This technique may be useful for scaling up our method to extremely large numbers of qubits.

### E. Setting the Error Parameters $\epsilon_1$ and $\epsilon_2$

Next, we describe how to set the parameters  $\epsilon_1$  and  $\epsilon_2$  in equations (III.9) and (III.10). We will use an approach that is similar to the one in Sections II B and II C.

First, we consider  $\epsilon_1$ , which bounds  $\mathbf{u}$ , the error in  $\mathbf{g}$ . Note that, when  $g_j$  is estimated using the procedure in Section II B, we also obtain large-deviation bounds on  $u_j$ . In particular, for some  $\delta_1 > 0$ , we have that:

$$\|u_j\|_{\psi_2} \leq \delta_1 c_{jj}, \quad (\text{III.19})$$

where  $\|\cdot\|_{\psi_2}$  is the sub-Gaussian norm (in the sense of [27]). (This bound can be obtained from equation (II.10), by setting  $N_{\text{trials}} \sim 1/\delta_1^2$ . Here, we are neglecting to count those trials of the experiment that are used to choose the evolution time  $t$ , because the number of those trials grows only logarithmically with  $\Gamma_{\mathbf{ab}}$ .)

This implies that  $\|\mathbf{u}\|_2^2$  is a subexponential random variable, whose subexponential norm (in the sense of [27]) is at most  $2\delta_1^2 \|\text{diag}(C)\|_2^2$ . This implies that  $\|\mathbf{u}\|_2$  is bounded with high probability: for any  $\tau \geq 1$ ,

$$\Pr[\|\mathbf{u}\|_2 \geq \tau\delta_1 \|\text{diag}(C)\|_2] \leq e \cdot \exp(-\tau^2/2c), \quad (\text{III.20})$$

where  $c > 0$  is some universal constant. We then choose  $\tau$  to be a sufficiently large constant, so that the failure probability is small. (Note that in some cases, one can prove stronger bounds, by taking advantage of the fact that the coordinates of  $\mathbf{u}$  are

independent, and using a Bernstein-type inequality [27]. This bound is stronger when the diagonal elements of  $C$  satisfy  $\|\text{diag}(C)\|_\infty \ll \|\text{diag}(C)\|_2$ , i.e., when  $\mathbf{u}$  has many independent coordinates with similar magnitudes.)

The above bound does not immediately tell us how to set  $\epsilon_1$ , because the bound depends on  $\text{diag}(C)$ , which is not known exactly. Instead, we now derive a bound that depends on  $\mathbf{g}$ , which is known explicitly, and can be used to set  $\epsilon_1$ . To do this, we assume that  $\delta_1$  is sufficiently small so that  $\tau\delta_1 < 1/4$ . With high probability, we have

$$\begin{aligned} \|\mathbf{u}\|_2 &\leq \tau\delta_1 \|\text{diag}(C)\|_2 \\ &\leq \tau\delta_1 (\|\mathbf{g}\|_2 + \|\mathbf{u}\|_2) \\ &\leq \frac{\tau\delta_1}{1 - \tau\delta_1} \|\mathbf{g}\|_2 =: \epsilon_1, \end{aligned} \quad (\text{III.21})$$

where we used the triangle inequality, and some algebra. This tells us how to set  $\epsilon_1$  so that equation (III.9) holds.

We remark that  $\epsilon_1$  can also be bounded in terms of  $\text{diag}(C)$ , as follows:

$$\begin{aligned} \epsilon_1 &\leq \frac{\tau\delta_1}{1 - \tau\delta_1} (\|\text{diag}(C)\|_2 + \|\mathbf{u}\|_2) \\ &\leq \frac{\tau\delta_1}{1 - \tau\delta_1} (1 + \tau\delta_1) \|\text{diag}(C)\|_2 \\ &\leq 2\tau\delta_1 \|\text{diag}(C)\|_2. \end{aligned} \quad (\text{III.22})$$

We will use this bound in Section V, when we analyze the accuracy of our estimate of  $C$ .

Next, we consider  $\epsilon_2$ , which bounds  $\mathbf{v}$ , the error in  $\mathbf{h}$ . We use the same approach as above. When  $h_j$  is estimated using the procedure in Section II B, we obtain a bound on the sub-Gaussian norm of  $v_j$ : for some  $\delta_2 > 0$ ,

$$\|v_j\|_{\psi_2} \leq \delta_2 \Phi_j(C). \quad (\text{III.23})$$

(This bound can be obtained from equation (II.10), by setting  $N_{\text{trials}} \sim 1/\delta_2^2$ . We allow  $\delta_2$  to be different from  $\delta_1$ , because the measurements used to estimate  $h_j$  are more costly than the measurements used to estimate  $g_j$ , hence one may prefer to use different values for  $N_{\text{trials}}$  in each case.)

This implies that  $\|\mathbf{v}\|_2^2$  is a subexponential random variable, hence  $\|\mathbf{v}\|_2$  is bounded with high probability: for any  $\tau \geq 1$ ,

$$\Pr[\|\mathbf{v}\|_2 \geq \tau\delta_2 \|\Phi(C)\|_2] \leq e \cdot \exp(-\tau^2/2c), \quad (\text{III.24})$$

where  $c > 0$  is some universal constant. We then choose  $\tau$  to be a sufficiently large constant, so that

the failure probability is small. (Note that, for typical choices of  $\Phi(\cdot)$ , we expect that  $\|\Phi(C)\|_\infty \ll \|\Phi(C)\|_2$ . This implies that  $\mathbf{v}$  has many independent coordinates with similar magnitudes. When this occurs, one can prove a stronger bound using a Bernstein-type inequality [27]. For simplicity, we do not use this more elaborate bound here.)

The above bound does not immediately tell us how to set  $\epsilon_2$ , because the bound depends on  $\Phi(C)$ , which is not known exactly. Instead, we now derive a bound that depends on  $\mathbf{h}$ , which is known explicitly, and can be used to set  $\epsilon_2$ . To do this, we assume that  $\delta_2$  is sufficiently small so that  $\tau\delta_2 < 1/4$ . With high probability, we have

$$\begin{aligned} \|\mathbf{v}\|_2 &\leq \tau\delta_2\|\Phi(C)\|_2 \\ &\leq \tau\delta_2(\|\mathbf{h}\|_2 + \|\mathbf{v}\|_2) \\ &\leq \frac{\tau\delta_2}{1 - \tau\delta_2}\|\mathbf{h}\|_2 =: \epsilon_2. \end{aligned} \quad (\text{III.25})$$

This tells us how to set  $\epsilon_2$  so that equation (III.10) holds.

Finally, we remark that  $\epsilon_2$  can also be bounded in terms of  $\|C\|_{\ell_1}$ , as follows:

$$\begin{aligned} \epsilon_2 &\leq \frac{\tau\delta_2}{1 - \tau\delta_2}(\|\Phi(C)\|_2 + \|\mathbf{v}\|_2) \\ &\leq \frac{\tau\delta_2}{1 - \tau\delta_2}(1 + \tau\delta_2)\|\Phi(C)\|_2 \\ &\leq 2\tau\delta_2\|\Phi(C)\|_2 \\ &\leq 2\tau\delta_2\sqrt{m}\|\Phi(C)\|_\infty \\ &\leq 4\tau\delta_2\sqrt{m}\|C\|_{\ell_1}. \end{aligned} \quad (\text{III.26})$$

We will use this bound in Section V, when we analyze the accuracy of our estimate of  $C$ .

#### IV. NUMERICAL EXAMPLES

We use numerical simulations to test how well our method performs on realistic system sizes, with different levels of sparsity, and when the data contain statistical fluctuations due to finite sample sizes. We find that our method performs well overall (see Figure 1.2).

We numerically simulate the protocol for randomly chosen  $C$  matrices (see Appendix B for details). In these examples we assume that the diagonal elements of  $C$  are known, that is,  $\epsilon_1 = 0$  in Eq. (III.11). We then solve the convex optimization problem given by (III.5)-(III.7) using CVXPY, a convex optimization package for Python [47, 48].

We first investigate the case of noiseless measurements, corresponding to  $\epsilon_2 = 0$  in Eq. (III.12). In Fig. 1.2(a) we show the recovery error as a function of the number of measurements,  $m$ , for a fixed number of qubits,  $n$ , and various choices of the off-diagonal sparsity,  $s$ . The sharp transition in the recovery error as a function of  $m$  is evident. Moreover, as shown in the inset of Fig. 1.2(a), the transition point  $m_c$ , which we define as the point where  $\|C - W^{(opt)}\|_\infty$  drops below 0.25,

scales linearly with  $s$ , consistent with our analytical results. In Fig. 1.2(b) we fix  $s$ , vary  $n$ , and study the recovery error as a function of  $m$ . Again, we observe a phase transition as  $m$  increases. In this case,  $m_c$  scales polynomially with  $\log(n)$  as suggested in the inset of Fig. 1.2(b).

We then investigate the effect of noisy measurements on the recovery error. We generate random  $C$  matrices, with a fixed number of qubits  $n$  and sparsity  $s$ . We simulate noise by adding a random vector  $\mathbf{e}$ , whose entries are independent Gaussian random variables with mean 0 and standard deviation  $\sigma_\epsilon$ , to measurement vector  $\mathbf{h}$ . We now replace (III.7) in the previous convex program with (III.12) and choose  $\epsilon_2 = \sqrt{m}\sigma_\epsilon$ . The scaling of the reconstruction error  $\|W^{(opt)} - C\|_\infty$  as a function of  $\sigma_\epsilon$  is shown in Fig. 1.2 (c). The recovery error after the phase transition point scales linearly with  $\sigma_\epsilon$ , consistent with our analytical bounds.

#### V. RECOVERY GUARANTEES

In this section we will study the convex optimization problem (III.13)-(III.16), and prove rigorous recovery guarantees that show that the optimal solution  $W^{(opt)}$  is close to the true correlation matrix  $C$ , provided that  $m \geq c_0 s \log n$  (and with better robustness to noise, when  $m \geq c_0 s \log^4 n$ ). Here,  $m$  is the dimension of the measurement vector  $\mathbf{h}$ ,  $s$  is the sparsity (the number of nonzero elements) in the off-diagonal part of the matrix  $C$ , and  $c_0$  is some universal constant.

Actually, we will prove two different results: a non-universal recovery guarantee, using the ‘‘RIPless’’ framework of [32], as well as a universal recovery guarantee, using RIP-based techniques [27–31, 33]. Here, RIP refers to the ‘‘restricted isometry property,’’ a fundamental proof technique in compressed sensing. There are different advantages to the RIPless and RIP-based bounds: the RIPless bounds require slightly fewer measurements, while the RIP-based bounds are more robust when the measurements are noisy.

Along the way, we will introduce two variants of the problem (III.13)-(III.16): constrained  $\ell_1$ -minimization and the LASSO (“least absolute shrinkage and selection operator”). Generally speaking, recovery guarantees that hold for one of these problems can be adapted to the other one, with minor modifications. Here, we follow [32] and prove a RIPless bound for the LASSO, and we follow [30, 33] and prove a RIP-based bound for constrained  $\ell_1$ -minimization.

### A. Simplifying the Problem

We start with the convex optimization problem (III.13)-(III.16). We first remove the positivity constraint  $d(W, K_+) \leq \epsilon_1$ ; this change should only hurt the accuracy of the solution  $W^{(opt)}$ . We also change the objective function to sum over all  $i < j$  rather than all  $i \neq j$ ; since  $W$  is symmetric, this merely changes the objective function by a factor of 2. Finally, we shift the variable  $W$  by subtracting away  $\text{diag}(\mathbf{g})$ , so that its diagonal elements are all zero. In similar way, we shift the measurement vector  $\mathbf{h}$  to get

$$\mathbf{h}' = \mathbf{h} - \Phi(\text{diag}(\mathbf{g})). \quad (\text{V.1})$$

This gives us an equivalent problem:

$$\text{Find } W \in \mathbb{R}^{n \times n} \text{ that minimizes } \sum_{i < j} |W_{ij}|, \quad (\text{V.2})$$

such that:

$$\text{diag}(W) = \mathbf{0}, \quad (\text{V.3})$$

$$\|\Phi(W) - \mathbf{h}'\|_2 \leq \epsilon_2 + \epsilon_1 \sqrt{mn}, \quad (\text{V.4})$$

$$W = W^T. \quad (\text{V.5})$$

We will use the following notation. We define an operation  $\text{diag}(\cdot)$  that has two meanings: given an  $n \times n$  matrix  $M$ ,  $\text{diag}(M)$  returns an  $n$ -dimensional vector containing the diagonal elements of  $M$ ; and given an  $n$ -dimensional vector  $\mathbf{v}$ ,  $\text{diag}(\mathbf{v})$  returns an  $n \times n$  matrix that contains  $\mathbf{v}$  along the diagonal, and zeroes off the diagonal.

Let us define  $C'$  to be the off-diagonal part of the correlation matrix  $C$ , that is,  $C'$  is the matrix whose off-diagonal elements match those of  $C$ , and whose diagonal elements are zero. We can write this concisely as:

$$C' = C - \text{diag}(\text{diag}(C)). \quad (\text{V.6})$$

We can view  $\mathbf{h}'$  as a measurement of  $C'$ , with additive error  $\mathbf{z}$ ,

$$\mathbf{h}' = \Phi(C') + \mathbf{z}. \quad (\text{V.7})$$

We want to show that the solution  $W^{(opt)}$  is an accurate estimate of  $C'$ . Note that we can write the error term  $\mathbf{z}$  in the form

$$\begin{aligned} \mathbf{z} &= \mathbf{h}' - \Phi(C') \\ &= \mathbf{h} - \Phi(C) - \Phi(\text{diag}(\mathbf{g} - \text{diag}(C))) \\ &= \mathbf{v} - \Phi(\text{diag}(\mathbf{u})), \end{aligned} \quad (\text{V.8})$$

where  $\mathbf{u}$  and  $\mathbf{v}$  are the noise terms in (III.9) and (III.10). Then we can bound  $\mathbf{z}$  using (III.10) and (III.18),

$$\|\mathbf{z}\|_2 \leq \epsilon_2 + \epsilon_1 \sqrt{mn}. \quad (\text{V.9})$$

It will be convenient to write

$$D = \{W \in \mathbb{R}^{n \times n} \mid W^T = W, \text{diag}(W) = \mathbf{0}\} \quad (\text{V.10})$$

to denote the subspace of symmetric matrices whose diagonal elements are all 0. Let  $\text{uvec} : D \rightarrow \mathbb{R}^{n(n-1)/2}$  denote the linear operator that returns the upper-triangular part of  $W$ ,

$$\text{uvec} : W \mapsto (W_{ij})_{i < j}. \quad (\text{V.11})$$

Let us write  $\Phi_D : D \rightarrow \mathbb{R}^m$  to denote the measurement operator  $\Phi$  restricted to act on the subspace  $D$  (this definition will be useful later, when we work with the adjoint operator  $\Phi_D^\dagger$ ). Then we can rewrite our problem (V.2)-(V.5) in a more concise form:

$$\begin{aligned} \text{Find } W \in D \text{ that minimizes } \|\text{uvec}(W)\|_1, \\ \text{such that:} \end{aligned} \quad (\text{V.12})$$

$$\|\Phi_D(W) - \mathbf{h}'\|_2 \leq \epsilon_2 + \epsilon_1 \sqrt{mn}. \quad (\text{V.13})$$

### B. LASSO Formulation

In the following discussion, we will also consider a variant of our problem, the LASSO [49]:

Find  $W \in D$  that minimizes:

$$\frac{1}{2} \|\Phi_D(W) - \mathbf{h}'\|_2^2 + \lambda \|\text{uvec}(W)\|_1. \quad (\text{V.14})$$

This can be viewed as a Lagrangian relaxation of the previous problem ((V.12)-(V.13)), or as an  $\ell_1$ -regularized least-squares problem. In addition, the convex optimization problems that were described earlier in Section III C can also be relaxed into a LASSO-like form, in a similar way. The choice of the regularization parameter  $\lambda$  requires some care. We will discuss this next.

### C. Setting the Regularization Parameter $\lambda$

In general, the regularization parameter  $\lambda$  controls the relative strength of the two parts of the objective function in (V.14). When the noise in the measurement of  $\mathbf{h}'$  is strong, then  $\lambda$  must be set large enough to ensure that the  $\ell_1$  regularization term still has the desired effect. However, if  $\lambda$  is too large, it strongly biases the solution  $W^{(opt)}$ , making it less accurate.

Here, we sketch one approach to setting  $\lambda$ , following the analysis in [32]. Our goal is to ensure that the solution  $W^{(opt)}$  converges to (a sparse approximation of) the true correlation matrix  $C$ . To do this, we must set  $\lambda$  large enough to satisfy two constraints, which involve the noise in the measurement of  $\mathbf{h}'$  (see equation (IV.1) and the equation below (IV.2) in [32]). When these constraints are satisfied, the error in the solution  $W^{(opt)}$  is bounded by equation (IV.2) in [32]. (Note that this error bound grows with  $\lambda$ , hence one should choose the smallest value of  $\lambda$  that satisfies the above constraints.)

We now show in detail how to carry out the above calculation, in order to set  $\lambda$ . First, we give precise statements of the two constraints on  $\lambda$ :

$$\|\text{uvec}(\Phi_D^\dagger \mathbf{z})\|_\infty \leq \lambda, \quad (\text{V.15})$$

$$\|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{z})\|_\infty \leq \lambda. \quad (\text{V.16})$$

Here,  $\mathbf{z}$  is the noise term in the measurement of  $\mathbf{h}'$  in equation (V.7);  $\Phi_D^\dagger : \mathbb{R}^m \rightarrow D$  is the adjoint of the measurement operator  $\Phi_D$ ;  $T \subset \{(j, j') \mid 1 \leq j < j' \leq n\}$  is the support of (a sparse approximation of) the true correlation matrix  $C$ ;  $\Phi_{D,T}$  is the sub-matrix of  $\Phi_D$  that contains those columns of  $\Phi_D$  whose indices belong to the set  $T$ ;  $P$  is the projection onto the range of  $\Phi_{D,T}$ ;  $T^c$  is the complement of the set  $T$ ; and  $\Phi_{D,T^c}$  is the sub-matrix of  $\Phi_D$  that contains those columns of  $\Phi_D$  whose indices belong to the set  $T^c$ .

In order to set  $\lambda$ , we need to compute the quantities in equations (V.15) and (V.16), and to do this, we need to have some bounds on the noise  $\mathbf{z}$ . We now demonstrate two ways of obtaining such bounds.

One straightforward way is as follows. We can use equation (V.8) to write

$$\mathbf{z} = \mathbf{v} - \Phi(\text{diag}(\mathbf{u})), \quad (\text{V.17})$$

where  $\mathbf{u}$  and  $\mathbf{v}$  are the noise terms in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$ , respectively. Also, recall that we previously showed bounds on  $\mathbf{u}$  and  $\mathbf{v}$  in Section III E, see equations (III.21) and (III.25). These imply bounds on  $\mathbf{z}$ , via an elementary calculation.

However, one can get better bounds on  $\mathbf{z}$  by using a more sophisticated approach, starting with bounds on the sub-Gaussian norms of  $u_j$  and  $v_j$ , such as equations (III.19) and (III.23). We describe this latter approach in detail.

We assume that  $\mathbf{g}$  and  $\mathbf{h}$  are measured using the procedures described in Sections II B and III E. Then equations (III.19) and (III.23) give us bounds on the sub-Gaussian norms of  $u_j$  and  $v_j$ :

$$\|u_j\|_{\psi_2} \leq \delta_1 c_{jj}, \quad (\text{V.18})$$

$$\|v_j\|_{\psi_2} \leq \delta_2 \Phi_j(C). \quad (\text{V.19})$$

Using these bounds, we can then set  $\lambda$  so that it satisfies (V.15) and (V.16) with high probability. More precisely, let us set

$$\lambda := (\epsilon_1''' \cdot 4\sqrt{mn} + \epsilon_2''')4\sqrt{m}(1 + \sqrt{2})\sqrt{\ln(n)/c'}, \quad (\text{V.20})$$

where

$$\epsilon_1''' := \frac{\delta_1}{1 - \epsilon_1''} \|\mathbf{g}\|_\infty, \quad \epsilon_1'' := 2\sqrt{\ln(n)/c_0} \delta_1, \quad (\text{V.21})$$

$$\epsilon_2''' := \frac{\delta_2}{1 - \epsilon_2''} \|\mathbf{h}\|_\infty, \quad \epsilon_2'' := 2\sqrt{\ln(m)/c_0} \delta_2. \quad (\text{V.22})$$

Here we are assuming that  $\delta_1$  and  $\delta_2$  are sufficiently small (for instance,  $\delta_1 \lesssim 1/\sqrt{\ln(n)}$  and  $\delta_2 \lesssim 1/\sqrt{\ln(n)}$ ) to ensure that  $\epsilon_1'' < 1/2$  and  $\epsilon_2'' < 1/2$ . Also, here  $c'$  and  $c_0$  are universal constants (which are defined in the proof below).

Now let the correlation matrix  $C$  and measurement operator  $\Phi_D$  be fixed, and note that the noise term  $\mathbf{z}$  is stochastic. Then we claim that, with high probability (over the random realization of  $\mathbf{z}$ ), equations (V.15) and (V.16) will be satisfied; here the failure probability is at most  $(e/n^3) + (e/m^3) + (2e/n^{1+2\sqrt{2}})$ . We prove this claim in Appendix C.

Finally, we have the following simple upper bounds on  $\epsilon_1'''$ ,  $\epsilon_2'''$  and  $\lambda$  (these follow from the definitions of  $\epsilon_1'''$  and  $\epsilon_2'''$ , the definitions of  $\mathbf{g}$  and  $\mathbf{h}$ , equations (C.4) and (C.7), and the definition of  $\Phi(\cdot)$ ):

$$\epsilon_1''' \leq 3\delta_1 \|\text{diag}(C)\|_\infty, \quad (\text{V.23})$$

$$\epsilon_2''' \leq 6\delta_2 \|C\|_{\ell_1}, \quad (\text{V.24})$$

$$\lambda \leq O(\sqrt{m \ln(n)}(\delta_1 \|\text{diag}(C)\|_\infty \sqrt{mn} + \delta_2 \|C\|_{\ell_1})). \quad (\text{V.25})$$

We will make use of these bounds later, when we analyze the accuracy of our estimate of  $C$ .



#### D. Isotropy and Incoherence of the Measurement Operator

We will show that the rows of the measurement operator  $\Phi_D$  have two properties, *isotropy* and *incoherence*, which play a fundamental role in compressed sensing (see, e.g., [27, 32]). Let  $Q$  be the matrix (of size  $m$  by  $n(n-1)/2$ ) that represents the action of  $\Phi_D$  (using the fact that the subspace  $D$  is isomorphic to  $\mathbb{R}^{n(n-1)/2}$ ); that is,  $Q$  and  $\Phi_D$  are related by the equation:

$$\Phi_D(C) = Q \cdot \text{uvec}(C), \quad \forall C \in D. \quad (\text{V.26})$$

The rows of  $Q$  are chosen independently at random, and each row has the form

$$\mathbf{q} = 4 \text{uvec}(\mathbf{r}\mathbf{r}^T) \in \mathbb{R}^{n(n-1)/2}, \quad (\text{V.27})$$

where  $\mathbf{r}$  is sampled from the distribution described in (III.2). We say that  $\mathbf{q}$  is *centered* if it has mean  $\mathbb{E}(\mathbf{q}) = \mathbf{0}$ , and we say that  $\mathbf{q}$  is *isotropic* if its covariance matrix is the identity:

$$\mathbb{E}(\mathbf{q}\mathbf{q}^T) = I. \quad (\text{V.28})$$

It is straightforward to check that  $\mathbf{q}$  is centered and isotropic (up to a normalization factor of 2), since:

$$\begin{cases} \mathbb{E}[r_i r_j] = 0 & i < j \\ \mathbb{E}[r_i r_j r_k r_l] = 0 & i < j \text{ and } k < l \text{ and } \{i, j\} \cap \{k, l\} = \emptyset \\ \mathbb{E}[r_i r_j r_k r_l] = 0 & i < j \text{ and } k < l \text{ and } |\{i, j\} \cap \{k, l\}| = 1 \\ \mathbb{E}[r_i r_j r_k r_l] = \frac{1}{4} & i < j \text{ and } k < l \text{ and } i = k \text{ and } j = l \end{cases} \quad (\text{V.29})$$

(Note that in the last line of Eq. V.29, we cannot have a case with  $i = l$  and  $j = k$ , as the requirements of  $i < j$  and  $k < l$  lead to a contradiction.)

In addition, we say that  $\mathbf{q}$  is *incoherent* with parameter  $\mu > 0$  if, with probability 1, all of its coordinates are small:

$$\|\mathbf{q}\|_\infty^2 \leq \mu. \quad (\text{V.30})$$

In order for this to be useful for compressed sensing, one needs  $\mu$  to be small, say, at most polylogarithmic in the dimension of  $\mathbf{q}$ . In our case, it is easy to see that  $\mathbf{q}$  is incoherent with parameter  $\mu = 16$ .

#### E. Non-universal (RIPless) Recovery Guarantee

We begin by proving a non-universal recovery guarantee, using the ‘‘RIPless’’ framework of [32], which in turn relies on the isotropy and incoherence properties shown in the preceding section.

Let  $C \in \mathbb{R}^{n \times n}$  be a correlation matrix, and let  $C' \in D$  be its off-diagonal part (see equation (V.6)). We will assume that  $C'$  is approximately  $s$ -sparse, i.e., there exists a matrix  $C^{(s)} \in D$  that has at most  $s$  nonzero entries above the diagonal, and that approximates  $C'$  in the (vector)  $\ell_1$  norm, up to an error of size  $\eta_s$ . This can be written compactly as:

$$\|\text{uvec}(C' - C^{(s)})\|_1 \leq \eta_s, \quad (\text{V.31})$$

where  $\text{uvec}(\cdot)$  was defined in equation (V.11). (Recall that both  $C'$  and  $C^{(s)}$  are symmetric, with all zeroes on the diagonal. Hence it suffices to consider those matrix entries that lie above the diagonal.)

We now choose the measurement operator  $\Phi$  at random (see equation (III.4)). We assume that  $m$  (the dimension of  $\mathbf{h}$ ) satisfies the bound:

$$m \geq \tilde{c}_0(1 + \beta) \cdot 4s \log \frac{n(n-1)}{2}. \quad (\text{V.32})$$

Here,  $\tilde{c}_0$  is a universal constant, and  $\beta > 0$  is a parameter that can be chosen freely by the experimenter. Note that  $m$  scales linearly with the sparsity  $s$ , but only logarithmically with the dimension  $n$  of the matrix  $C$ . This scaling is close to optimal, in an information-theoretic sense. This is one way of quantifying the advantage of compressed sensing, when compared with measurements that do not exploit the sparsity of  $C$ .

We measure  $\mathbf{g} \approx \text{diag}(C)$  and  $\mathbf{h} \approx \Phi(C)$ , and we calculate  $\mathbf{h}'$  (see equation (V.1)). We let  $\delta_1$  and  $\delta_2$  quantify the noise in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$ , as described in equations (V.18) and (V.19). We then solve the LASSO problem in (V.14), setting the regularization parameter  $\lambda$  according to (V.20). Let  $W^{(opt)}$  be the solution of this optimization problem.

We now have the following recovery guarantee, which shows that  $W^{(opt)}$  gives a good approximation to  $C'$ , in both the Frobenius ( $\ell_2$ ) norm, and the vector  $\ell_1$  norm. This follows directly from Theorem 1.2 in [32] (and the extension of that theorem to more general classes of noisy measurements in Section IV in [32]).

**Theorem V.1** *For any correlation matrix  $C$  satisfying the sparsity condition (V.31), with probability at least  $1 - \frac{12}{n(n-1)} - 6e^{-\beta}$  (over the random choice of the measurement operator  $\Phi$ , with  $m$  set according*

to (V.32)), the solution  $W^{(opt)}$  is close to  $C'$ , with an error that is bounded by:

$$\|W^{(opt)} - C'\|_F \leq \sqrt{2}c(1 + \alpha) \left[ \frac{\eta_s}{\sqrt{s}} + \lambda\sqrt{s} \right], \quad (\text{V.33})$$

$$\|W^{(opt)} - C'\|_{\ell_1} \leq 2c(1 + \alpha)[\eta_s + \lambda s], \quad (\text{V.34})$$

where  $c$  is a universal constant, and  $\alpha \leq \log^{3/2}(\frac{n(n-1)}{2})$ .

In these bounds, the first term upper-bounds the error that results from approximating  $C'$  by a sparse

matrix, and the second term upper-bounds the error due to noise in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$ .

To make the second term more transparent, we can combine it with the bound on  $\lambda$  from equation (V.25):

$$\lambda \leq O(\sqrt{m \ln(n)}(\delta_1 \|\text{diag}(C)\|_\infty \sqrt{mn} + \delta_2 \|C\|_{\ell_1})), \quad (\text{V.35})$$

where  $\delta_1$  and  $\delta_2$  quantify the noise in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$ , as described earlier.

Also, it is useful to consider the special case where  $C'$  is *exactly*  $s$ -sparse, so  $\eta_s = 0$ , and where we use as few measurement settings as possible, by setting  $m = O(s \log n)$ . In this case, we have:

$$\|W^{(opt)} - C'\|_F \leq O(\log^{3/2}(n)\lambda\sqrt{s}) \quad (\text{V.36})$$

$$\leq O(\sqrt{s} \log^{3/2}(n) \sqrt{m \log(n)} \cdot [\delta_1 \|\text{diag}(C)\|_\infty \sqrt{mn} + \delta_2 \|C\|_{\ell_1}]) \quad (\text{V.37})$$

$$\leq O(s \log^{5/2}(n) [\delta_1 \|\text{diag}(C)\|_\infty \sqrt{sn \log(n)} + \delta_2 \|C\|_{\ell_1}]) \quad (\text{V.38})$$

$$\leq O(s \log^{5/2}(n) [\delta_1 \|\text{diag}(C)\|_\infty \sqrt{sn \log(n)} + \delta_2 \sqrt{n} \|\text{diag}(C)\|_2 + \delta_2 \sqrt{2s} \|C'\|_F]). \quad (\text{V.39})$$

This can be compared with the error bound (II.31) for the naive method, and the RIP-based error bound (V.48) for compressed sensing. Generally speaking, compressed sensing has an advantage over the naive method when  $s$  is small, and the RIP-less bound is useful in the regime between  $m = O(s \log n)$  and  $m = O(s \log^4 n)$ , where the RIP-based bound does not apply. (When  $m$  is  $O(s \log^4 n)$  or larger, the RIP-based bound applies, and gives better results than the RIPless bound.) We will carry out a more detailed comparison between the naive method and compressed sensing in Section VI.

## F. Universal (RIP-based) Recovery Guarantee

Next, we prove a universal recovery guarantee, using an older approach based on the restricted isometry property (RIP) [27–31, 33]. This also relies on the isotropy and incoherence properties shown above. As these techniques are fairly standard in compressed sensing, we will simply sketch the proof.

First, we set the number of measurement settings to be

$$m \geq c_0 s \log^4 n, \quad (\text{V.40})$$

where  $s$  is the sparsity parameter, and  $c_0$  is some universal constant. (Note that  $m$  is slightly larger, by

a poly( $\log n$ ) factor, compared to the RIPless case.) Also, recall that  $D$  is the subspace of symmetric matrices whose diagonal elements are all 0 (see equation (V.10)), and  $\Phi_D$  is the measurement operator restricted to act on this subspace.

We claim that, with high probability (over the random choice of  $\Phi_D$ ), the normalized measurement operator  $\Phi_D/\sqrt{m}$  satisfies the RIP (for sparsity level  $2s$ ). To see this, we recall the isotropy and incoherence properties shown above. These properties imply that the measurement operator  $\Phi_D$  is sampling at random from a “bounded orthonormal system.” Such operators are known to satisfy the RIP, via a highly nontrivial proof [28, 29]; a recent exposition can be found in Chapter 12 in [30].

From this point onwards, we let the measurement operator  $\Phi_D$  be fixed. We will show that  $\Phi_D$  is capable of reconstructing the off-diagonal parts of *all* sparse matrices  $C$ , i.e.,  $\Phi_D$  can perform “universal” recovery.

As in the previous section, let  $C \in \mathbb{R}^{n \times n}$  be a correlation matrix, and let  $C' \in D$  be its off-diagonal part (see equation (V.6)). We will assume that  $C'$  is approximately  $s$ -sparse, i.e., there exists a matrix  $C^{(s)} \in D$  that has at most  $s$  nonzero entries above the diagonal, and that approximates  $C'$  in the (vector)  $\ell_1$  norm, up to an error of size  $\eta_s$ . This can be

written compactly as:

$$\|\text{uvec}(C' - C^{(s)})\|_1 \leq \eta_s, \quad (\text{V.41})$$

where  $\text{uvec}(\cdot)$  was defined in equation (V.11). (Recall that both  $C'$  and  $C^{(s)}$  are symmetric, with all zeroes on the diagonal. Hence it suffices to consider those matrix entries that lie above the diagonal.)

We measure  $\mathbf{g} \approx \text{diag}(C)$  and  $\mathbf{h} \approx \Phi(C)$ , and we calculate  $\mathbf{h}'$  (see equation (V.1)). We assume that the noise in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$  is bounded by  $\delta_1$  and  $\delta_2$ , as described in (III.19) and (III.23). We then solve the  $\ell_1$ -minimization problem in (V.12) and (V.13), setting the parameters  $\epsilon_1$  and  $\epsilon_2$  according to (III.21) and (III.25). Let  $W^{(opt)}$  be the solution of this problem.

We now have the following recovery guarantee, which shows that  $W^{(opt)}$  gives a good approximation to  $C'$ , in the Frobenius ( $\ell_2$ ) norm, and in the  $\ell_1$  vector norm. This follows directly from Theorem 1.9 in [33], and Theorem 6.12 in [30]. (There is one subtle point: the convex optimization problem in (V.12) and (V.13) uses the unnormalized measurement operator  $\Phi_D$ , while the error bounds in [30, 33] apply to the normalized measurement operator  $\Phi_D/\sqrt{m}$ . Hence, the noise is smaller by a factor of  $\sqrt{m}$  in these error bounds.)

**Theorem V.2** *With high probability (over the choice of the measurement operator  $\Phi$ , with  $m$  set*

*according to (V.40)), for all correlation matrices  $C$  (satisfying the sparsity condition (V.41)), the solution  $W^{(opt)}$  satisfies the following error bounds:*

$$\|W^{(opt)} - C'\|_F \leq c_1 \frac{\eta_s}{\sqrt{s}} + c_2 \left( \frac{\epsilon_2}{\sqrt{m}} + \epsilon_1 \sqrt{n} \right), \quad (\text{V.42})$$

$$\|W^{(opt)} - C'\|_{\ell_1} \leq c_1 \eta_s + c_2 \sqrt{s} \left( \frac{\epsilon_2}{\sqrt{m}} + \epsilon_1 \sqrt{n} \right), \quad (\text{V.43})$$

where  $c_1$  and  $c_2$  are universal constants.

In these bounds, the first term upper-bounds the error that results from approximating  $C'$  by a sparse matrix, and the second term upper-bounds the error due to noise in the measurements of  $\mathbf{g}$  and  $\mathbf{h}$ .

In order to apply these bounds, one needs to know the values of  $\epsilon_1$  and  $\epsilon_2$ . These can be obtained from Section III E, equations (III.22) and (III.26):

$$\epsilon_1 \leq O(\delta_1 \|\text{diag}(C)\|_2), \quad (\text{V.44})$$

$$\epsilon_2 \leq O(\delta_2 \sqrt{m} \|C\|_{\ell_1}). \quad (\text{V.45})$$

Also, it is useful to consider the special case where  $C'$  is *exactly*  $s$ -sparse, so  $\eta_s = 0$ , and where we use as few measurement settings as possible, by setting  $m = O(s \log^4 n)$ . In this case, we have:

$$\|W^{(opt)} - C'\|_F \quad (\text{V.46})$$

$$\leq O(\sqrt{n} \delta_1 \|\text{diag}(C)\|_2 + \delta_2 \|C\|_{\ell_1}) \quad (\text{V.47})$$

$$\leq O(\sqrt{n} \delta_1 \|\text{diag}(C)\|_2 + \delta_2 (\sqrt{n} \|\text{diag}(C)\|_2 + \sqrt{2s} \|C'\|_F)). \quad (\text{V.48})$$

This can be compared with the error bound (II.31) for the naive method, and the RIPless error bound (V.39) for compressed sensing. Generally speaking, compressed sensing has an advantage over the naive method when  $s$  is small, and the RIP-based bound has better scaling (as a function of  $n$ ,  $s$ ,  $\text{diag}(C)$  and  $C'$ ) than the RIPless bound, although it requires  $m$  to be slightly larger. We will carry out a more detailed comparison between the naive method and compressed sensing in Section VI.

## VI. PERFORMANCE EVALUATION

In this section we study the performance of our compressed sensing method, for a typical measurement scenario. We consider both the accuracy of the method, and the experimental resources required to implement it. We investigate the asymptotic scaling of our method, and compare it to the naive method, direct estimation of the correlation matrix, introduced in Section II C.

Overall, we find that our compressed sensing method has asymptotically better sample complexity, whenever the off-diagonal part of the correlation matrix  $C$  is sufficiently sparse. In particular, for

a system of  $n$  qubits, our method is advantageous whenever the number of correlated pairs of qubits,  $s$ , is at most  $O(n^{3/2})$  (ignoring log factors). These results are summarized in Figure VI.1.

We now explain these results in detail. We let  $C'$  be the off-diagonal part of the correlation matrix  $C$ , that is,  $C'$  is the matrix whose off-diagonal elements match those of  $C$ , and whose diagonal elements are zero. We are promised that  $C'$  has at most  $s$  nonzero elements above the diagonal (and, by symmetry, at most  $s$  nonzero elements below the diagonal). Our goal is to estimate both  $C'$  and  $\text{diag}(C)$ .

Compressed sensing allows the possibility of adjusting the number of measurement settings,  $m$ , over a range from  $\sim s \log n$  to  $n^2$ . (Note that  $m \sim s \log n$  is just slightly above the information-theoretic lower bound, while  $m = n^2$  is the number of measurement settings used by the naive method.) Compressed sensing works across this whole range, but the error bounds vary depending on  $m$ . There are two cases: (1) For  $m \gtrsim s \log^4 n$ , both the RIP-based and RIP-less error bounds are available, and the RIP-based error bounds are asymptotically stronger. (2) For  $m$  between  $\sim s \log n$  and  $\sim s \log^4 n$ , only the RIPless error bound is available.

To make a fair comparison between compressed sensing and the naive method, we need to quantify the accuracy of these methods in a consistent way. This is a nontrivial task, because the error bounds for the different methods have different dependences on the parameters  $n$ ,  $s$ ,  $\text{diag}(C)$  and  $C'$ ; this can be seen by comparing equation (II.31), and Theorems V.1 and V.2.

We choose a simple way of quantifying the accuracy of all of these methods: given some  $\delta > 0$ , we require that each method return an estimate  $\hat{C}'$  that satisfies

$$\|\hat{C}' - C'\|_F \leq \delta \|C'\|_F. \quad (\text{VI.1})$$

Here, we use the Frobenius matrix norm, which is equivalent to the vector  $\ell_2$  norm. We write  $C$  (rather than  $C'$ ) on the right hand side of the inequality, in order to allow the recovery error to depend on both the diagonal and the off-diagonal elements of  $C$ .

In addition, we require that each method return an estimate  $\mathbf{g}$  of  $\text{diag}(C)$  that satisfies

$$\|\mathbf{g} - \text{diag}(C)\|_2 \leq \delta \|\text{diag}(C)\|_2. \quad (\text{VI.2})$$

For both compressed sensing as well as the naive method,  $\mathbf{g}$  is obtained in the same way, by performing single-qubit spectroscopy as in (II.19), and the error in  $\mathbf{g}$  satisfies the same bound (II.30).

We also need to account for the cost of implementing each method using real experiments. This cost depends on a number of factors. One factor is the total number of experiments that have to be performed, often called the *sample complexity*. This is the number of measurement settings, times the number of repetitions of the experiment with each measurement setting. Another factor is the difficulty of performing a single run of the experiment. This involves both the difficulty of preparing entangled states (random  $n$ -qubit GHZ states for the compressed sensing method, and 2-qubit Bell states for the naive method), and the length of time that one has to wait in order to observe dephasing.

Here, we study a scenario where we expect our compressed sensing method to perform well. We consider an advanced quantum information processor, where  $n$ -qubit GHZ states are fairly easy to prepare (using  $O(\log n)$ -depth quantum circuits), and dephasing occurs at low rates, so that the main cost of running each experiment is the amount of time needed to observe dephasing. In this scenario, it is reasonable to use the sample complexity as a rough measure of the total cost of implementing the compressed sensing method, as well as the naive method.

We now calculate the sample complexity for three methods of interest: (1) the naive method with  $m = n^2$ , (2) compressed sensing with  $m \sim s \log^4 n$ , and (3) compressed sensing with  $m \sim s \log n$ . We find that method (2) outperforms the naive method whenever  $s \leq O(n^{3/2}/\log^2 n)$ , and method (3) outperforms the naive method whenever  $s \leq O(n^{2/3}/\log^2 n)$ .

In addition, for each of these methods, we show the number of samples where single-qubit spectroscopy is performed, and the number of samples where multi-qubit spectroscopy is performed. (Recall that all of these methods use single-qubit spectroscopy to estimate the diagonal of  $C$ , and then use multi-qubit spectroscopy to estimate the off-diagonal part of  $C$ .) Both of these numbers can be important: multi-qubit spectroscopy is more expensive to implement on essentially all experimental platforms, and requires more samples when  $s$  is large; but it is possible for single-qubit spectroscopy to dominate the overall sample complexity, when  $s$  is small.

### A. Naive method with $m = n^2$

As in Section II C, we use two parameters,  $\delta_1$  and  $\delta_2$ , to quantify the accuracy of the measurements, as in equations (II.24) and (II.25). Then we get an es-

Reconstruction method:	Naive	CS (RIP-based)	CS (RIPless*)
<b>Single-qubit spectroscopy:</b>			
# of meas. settings	$n$	$n$	$n$
# of samples per setting	$O(n/\delta^2)$	$O(n/\delta^2)$	$O(s^3 \log^6(n)/\delta^2)$
Total # of samples	$O(n^2/\delta^2)$	$O(n^2/\delta^2)$	$O(ns^3 \log^6(n)/\delta^2)$
<b>Multi-qubit spectroscopy:</b>			
# of meas. settings $m$	$O(n^2)$	$O(s \log^4 n)$	$O(s \log n)$
# of samples per setting	$O(n/\delta^2)$	$O(\max(n, s)/\delta^2)$	$O(s^2 \max(n, s) \log^5(n)/\delta^2)$
Total # of samples	$O(n^3/\delta^2)$	$O(s \max(n, s) \log^4(n)/\delta^2)$	$O(s^3 \max(n, s) \log^6(n)/\delta^2)$
<b>Total sample complexity:</b>	$O(n^3/\delta^2)$	$O(\max(n, s)^2 \log^4(n)/\delta^2)$	$O(s^3 \max(n, s) \log^6(n)/\delta^2)$

FIG. VI.1. Sample complexity of different methods for reconstructing a correlation matrix  $C$ , of size  $n \times n$ , with  $2s$  nonzero elements off the diagonal. The naive method is to measure each element of  $C$  separately. “CS” refers to the compressed sensing method, and “RIP-based” and “RIPless” refer to different analytical bounds on the accuracy of the reconstruction of  $C$ . The asterisk (\*) indicates that the results using the RIPless bound hold under a technical assumption that the diagonal of  $C$  does not have any unusually large elements, see equation (VI.10). The different methods are parameterized in such a way that they reconstruct the diagonal of  $C$  up to an additive error of size  $\delta \|\text{diag}(C)\|_2$ , and they reconstruct the off-diagonal part of  $C$  up to an additive error of size  $\delta \|C\|_F$ . Each method makes use of single-qubit spectroscopy (with  $n$  experimental configurations or “measurement settings”), as well as multi-qubit spectroscopy (with  $m$  measurement settings, where  $m$  varies between  $O(n^2)$  and  $O(s \log n)$ ). The CS method has lower sample complexity than the naive method, when  $s \ll n^2$ . In particular, using the RIP-based bound with  $m = O(s \log^4 n)$ , the CS method is advantageous whenever  $s \leq O(n^{3/2}/\log^2 n)$ . Using the RIPless bound with  $m = O(s \log n)$ , the CS method is advantageous whenever  $s \leq O(n^{2/3}/\log^2 n)$ .

estimate  $\widehat{C}'$  of  $C'$ , whose error is bounded by equation (II.31): with probability at least  $1 - \eta$ ,

$$\begin{aligned} & \|\widehat{C}' - C'\|_F^2 \\ & \leq \frac{1}{\eta} [3(n-1)(\delta_1 + \delta_2)^2 \|\text{diag}(C)\|_2^2 + 6\delta_2^2 \|C'\|_F^2]. \end{aligned} \quad (\text{VI.3})$$

For simplicity, we set  $\eta$  to be some universal constant, say  $\eta = 0.001$ . Now, given any  $\delta > 0$ , we can ensure that

$$\begin{aligned} \|\widehat{C}' - C'\|_F^2 & \leq \delta^2 \|\text{diag}(C)\|_2^2 + (\delta^2/n) \|C'\|_F^2 \\ & \leq \delta^2 \|C\|_F^2, \end{aligned} \quad (\text{VI.4})$$

by setting  $\delta_1 = \delta_2 = O(\delta/\sqrt{n})$ . This satisfies the requirement (VI.1).

In addition, one can easily check that the estimate  $\mathbf{g}$  for  $\text{diag}(C)$  satisfies the requirement (VI.2).

Then the sample complexity is as follows (see the discussion preceding (II.24) and (II.25)): the method performs single-qubit spectroscopy on  $O(n/\delta_1^2) = O(n^2/\delta^2)$  samples, and multi-qubit spectroscopy on  $O(n^2/\delta_2^2) = O(n^3/\delta^2)$  samples. Hence the total sample complexity is  $O(n^3/\delta^2)$ .

#### B. Compressed sensing with $m \sim s \log^4 n$

Here we consider the  $\ell_1$ -minimization problem in (V.12) and (V.13), whose solution  $W^{(opt)}$  satisfies the RIP-based bound in Theorem V.2. We use two parameters,  $\delta_1$  and  $\delta_2$ , to quantify the accuracy of the measurements, as in equations (III.19) and (III.23). Section III E then explains how to set the parameters  $\epsilon_1$  and  $\epsilon_2$  that appear in (V.12) and (V.13).

The estimator  $W^{(opt)}$  satisfies the following error bound (see Theorem V.2, and equation (V.48)):

$$\|W^{(opt)} - C'\|_F \leq O(\sqrt{n}\delta_1 \|\text{diag}(C)\|_2 + \delta_2(\sqrt{n}\|\text{diag}(C)\|_2 + \sqrt{2s}\|C'\|_F)). \quad (\text{VI.5})$$

Now, given any  $\delta > 0$ , we can ensure that

$$\begin{aligned} \|W^{(opt)} - C'\|_F &\leq \frac{1}{\sqrt{2}}\delta\|\text{diag}(C)\|_2 + \frac{1}{\sqrt{2}}\delta\|C'\|_F \\ &\leq \delta\|C\|_F, \end{aligned} \quad (\text{VI.6})$$

by setting  $\delta_1 = O(\delta/\sqrt{n})$  and

$$\delta_2 = O(\delta/\sqrt{\max(n, s)}). \quad (\text{VI.7})$$

This satisfies the requirement (VI.1).

In addition, one can easily check that the estimate  $\mathbf{g}$  for  $\text{diag}(C)$  satisfies the requirement (VI.2).

Then the sample complexity is as follows (see the discussion following (III.19) and (III.23)): the method performs single-qubit spectroscopy on  $O(n/\delta_1^2) = O(n^2/\delta^2)$  samples, and multi-qubit spectroscopy on

$$O(m/\delta_2^2) \leq O(s \max(n, s) \log^4(n)/\delta^2) \quad (\text{VI.8})$$

samples. Hence the total sample complexity is at most

$$O(\max(n, s)^2 \log^4(n)/\delta^2). \quad (\text{VI.9})$$

This is less than the sample complexity of the naive method, provided the off-diagonal part of the correlation matrix is sufficiently sparse, i.e., when  $s \leq O(n^{3/2}/\log^2 n)$ .

### C. Compressed sensing with $m \sim s \log n$

Here we consider the LASSO optimization problem in (V.14), whose solution  $W^{(opt)}$  satisfies the RIPless bound in Theorem V.1. We use two parameters,  $\delta_1$  and  $\delta_2$ , to quantify the accuracy of the measurements, as in equations (III.19) and (III.23).

Section VC then explains how to set the LASSO regularization parameter  $\lambda$ .

In the following, we assume that the diagonal elements of  $C$  satisfy a bound of the form

$$\|\text{diag}(C)\|_\infty \leq O\left(\frac{1}{\sqrt{n}}\|\text{diag}(C)\|_2\right). \quad (\text{VI.10})$$

We will first discuss the situations when this assumption holds; then we will use this assumption to get a stronger error bound for  $W^{(opt)}$ .

Roughly speaking, the assumption (VI.10) says that none of the diagonal elements  $c_{jj}$  is too much larger than the others. This is plausible for a quantum system that consists of many qubits that are constructed in a similar way.

In order to make this intuition more precise, we can write (VI.10) in an equivalent form:

$$\max_{1 \leq j \leq n} (c_{jj}^2) \leq O\left(\frac{1}{n} \sum_{j=1}^n c_{jj}^2\right), \quad (\text{VI.11})$$

which says that the largest  $c_{jj}^2$  is at most a constant factor larger than the average of all of the  $c_{jj}^2$ . Also, it is informative to consider how (VI.10) and (VI.11) compare to the (arguably more natural) assumption that

$$\max_{1 \leq j \leq n} |c_{jj}| \leq O\left(\frac{1}{n} \sum_{j=1}^n |c_{jj}|\right). \quad (\text{VI.12})$$

In fact, (VI.12) is actually a stronger assumption, in the sense that it implies (VI.10) and (VI.11), via the Cauchy-Schwartz inequality.

The estimator  $W^{(opt)}$  satisfies an error bound given by Theorem V.1, and equation (V.39). Combining this with our assumption (VI.10), we get the following:

$$\|W^{(opt)} - C'\|_F \leq O(s \log^{5/2}(n) [\delta_1 \sqrt{s \log(n)} \|\text{diag}(C)\|_2 + \delta_2 \sqrt{n} \|\text{diag}(C)\|_2 + \delta_2 \sqrt{2s} \|C'\|_F]). \quad (\text{VI.13})$$

Now, given any  $\delta > 0$ , we can ensure that

$$\begin{aligned} \|W^{(opt)} - C'\|_F &\leq \frac{1}{\sqrt{2}}\delta\|\text{diag}(C)\|_2 + \frac{1}{\sqrt{2}}\delta\|C'\|_F \\ &\leq \delta\|C\|_F, \end{aligned} \quad (\text{VI.14})$$

by setting

$$\delta_1 = O\left(\frac{\delta}{s^{3/2} \log^3(n)}\right), \quad (\text{VI.15})$$

and

$$\delta_2 = O\left(\frac{\delta}{s \log^{5/2}(n) \sqrt{\max(n, s)}}\right). \quad (\text{VI.16})$$

This satisfies the requirement (VI.1).

In addition, one can easily check that the estimate  $\mathbf{g}$  for  $\text{diag}(C)$  satisfies the requirement (VI.2).

Then the sample complexity is as follows (see the discussion following (III.19) and (III.23)): the method performs single-qubit spectroscopy on

$$O(n/\delta_1^2) = O(ns^3 \log^6(n)/\delta^2) \quad (\text{VI.17})$$

samples, and multi-qubit spectroscopy on

$$O(m/\delta_2^2) \leq O(s^3 \max(n, s) \log^6(n)/\delta^2) \quad (\text{VI.18})$$

samples. Hence the total sample complexity is at most

$$O(s^3 \max(n, s) \log^6(n)/\delta^2). \quad (\text{VI.19})$$

This is less than the sample complexity of the naive method, provided the off-diagonal part of the correlation matrix is sufficiently sparse, i.e., when  $s \leq O(n^{2/3}/\log^2 n)$ .

## VII. CHOOSING THE EVOLUTION TIME $t$

We now discuss a technical detail involving the physical implementation of our measurements of the correlation matrix  $C$ . As described in Section II B, this requires estimating certain decay rates  $\Gamma_{\mathbf{ab}}$ . To do this, we prepare quantum states  $|\psi_{\mathbf{ab}}\rangle$ , allow them to evolve for some time  $t$ , and then measure them in an appropriate basis. This works well when  $t$  is chosen appropriately, so that  $\Gamma_{\mathbf{ab}}t \sim 1$ .

In this section, we sketch one way of choosing the evolution time  $t$  such that  $\frac{1}{2} \leq \Gamma_{\mathbf{ab}}t \leq 2$ . The basic idea is to start with some initial guess for  $t$  (call it  $\tau_0$ ), then perform “binary search,” i.e., run a sequence of experiments, where one observes the dephasing of the state  $|\psi_{\mathbf{ab}}\rangle$  for some time  $t$ , and after each experiment, one adjusts the time  $t$  adaptively, multiplying and dividing by factors of 2, in order to get the “right amount” of dephasing. We claim that this requires  $\sim |\log(\Gamma_{\mathbf{ab}}\tau_0)|$  experiments.

More precisely, we consider the following procedure:

1. Fix some  $\tau_0 > 0$ ; this is our initial guess for the evolution time  $t$ .

2. For  $r = 1, 2, \dots, N_{\text{trials}}$ , do the following: (we set  $N_{\text{trials}}$  according to equation (VII.9) below)

- (a) Set  $s_0 = 0$  and  $t_0 = 2^{s_0}\tau_0$ . (This is our initial guess for  $t$ .)

- (b) For  $j = 0, 1, 2, \dots, N_{\text{steps}} - 1$ , do the following: (we set  $N_{\text{steps}}$  according to equation (VII.5) below)

- i. Prepare the state  $|\psi_{\mathbf{ab}}\rangle = \frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ , allow the state to dephase for time  $t_j$ , then measure in the basis  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle \pm |\mathbf{b}\rangle)$

- ii. If the measurement returns  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle + |\mathbf{b}\rangle)$ , then set

$$s_{j+1} = \begin{cases} s_j + 1 & \text{with probability } \frac{e-1}{e+1}, \\ s_j & \text{otherwise.} \end{cases} \quad (\text{VII.1})$$

If the measurement returns  $\frac{1}{\sqrt{2}}(|\mathbf{a}\rangle - |\mathbf{b}\rangle)$ , then set  $s_{j+1} = s_j - 1$ .

- iii. Set  $t_{j+1} = 2^{s_{j+1}}\tau_0$ . (This is our next guess for  $t$ .)

- (c) Define  $\xi_r$  to be the value of  $s_j$  from the last iteration of the loop, that is,  $\xi_r = s_{N_{\text{steps}}}$ .

3. Compute the average  $\xi = \frac{1}{N_{\text{trials}}} \sum_{r=1}^{N_{\text{trials}}} \xi_r$ . Return  $\hat{t} = 2^\xi \tau_0$ . (This is our estimate for  $t$ .)

This procedure can be described in an intuitive way as follows. The inner loop of this procedure (the loop indexed by  $j$ ) can be viewed as a kind of stochastic gradient descent, which behaves like a random walk on real numbers of the form  $t = 2^s \tau_0$  ( $s \in \mathbb{Z}$ ) (see the dashed curves in Fig. VII.1(a)).

We will show that this random walk has a single basin of attraction at a point  $t^* = 2^{s^*} \tau_0$  that satisfies  $\Gamma_{\mathbf{ab}}t^* \approx 1$ , that is,  $s^* \approx -\log_2(\Gamma_{\mathbf{ab}}\tau_0)$ . We claim that the random walk converges to this point: with high probability, the sequence  $s_0, s_1, s_2, \dots$  will reach the point  $s^*$  after  $O(|s^*|) = O(|\log(\Gamma_{\mathbf{ab}}\tau_0)|)$  steps; after that point, the sequence will remain concentrated around  $s^*$ , with exponentially-decaying tail probabilities (see Fig. VII.1(b)). This claim is made precise in Section VII A, equations (VII.5) and (VII.6).

Finally, the outer loop of this procedure (the loop indexed by  $r$ ) computes an estimate  $\xi$  of  $s^*$ , by averaging over several independent trials (see the solid curves in Fig. VII.1(a)). This then yields an estimate  $\hat{t}$  of  $t^*$ . The required number of trials, and the

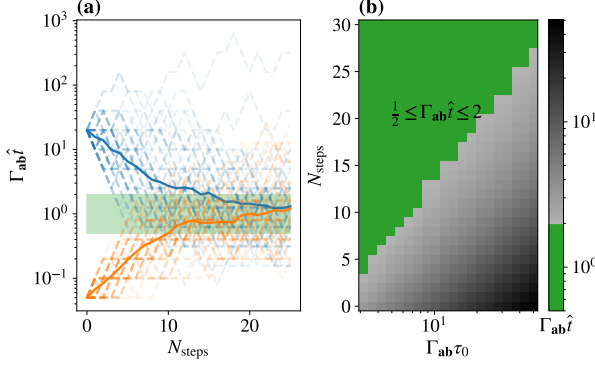


FIG. VII.1. Choosing the evolution time  $t$ . (a) Trajectories of the random walk. We start with an initial guess  $\tau_0$ , that can either be shorter (orange) or longer (blue) than the optimal time. The evolution time is then stochastically halved or doubled over  $N_{\text{steps}}$  iterations, according to the algorithm outlined in Section VII. This procedure is repeated (dashed curves) and the outcome is averaged (solid curves) to obtain an estimate  $\hat{t}$ . The region in which  $\frac{1}{2} < \Gamma_{ab}\hat{t} < 2$  is shaded in green. (b) The accuracy of the final estimate  $\hat{t}$ , as a function of the number of steps  $N_{\text{steps}}$  and the starting point  $\tau_0$ . The green shading shows the region where  $\hat{t}$  satisfies the bound  $\frac{1}{2} < \Gamma_{ab}\hat{t} < 2$ . We see that, at the boundary of the green region,  $N_{\text{steps}}$  scales logarithmically with  $|\Gamma_{ab}\tau_0|$ , as predicted by Eq. (VII.5).

accuracy of the resulting estimate  $\hat{t}$ , are analyzed in Section VII B, equations (VII.9) and (VII.10).

### A. Convergence of the Random Walk

We now give a rigorous analysis of our procedure for choosing  $t$ . We begin by describing the random walk in more detail. We will work with the variables  $s_j$ , which are related to the  $t_j$  via the identity  $t_j = 2^{s_j}\tau_0$ . It is easy to see that  $s_0 = 0$ ,  $s_{j+1} \in \{s_j, s_j - 1, s_j + 1\}$ , and

$$\begin{aligned} \mathbb{E}[s_{j+1} | s_j] &= s_j + \frac{1}{2}(1 + e^{-\Gamma_{ab}t_j})\frac{e-1}{e+1} - \frac{1}{2}(1 - e^{-\Gamma_{ab}t_j}) \\ &= s_j + \frac{1}{e+1}(e^{1-\Gamma_{ab}t_j} - 1). \end{aligned} \quad (\text{VII.2})$$

Hence the sequence  $s_0, s_1, s_2, \dots$  can be viewed as the trajectory of a random walk on a 1-dimensional chain, beginning at  $s_0$ , with transition probabilities that vary along the chain. The expected behavior of

the random walk can be bounded as follows:

$$\mathbb{E}[s_{j+1} | s_j] \begin{cases} \geq s_j + \mu & \text{when } \Gamma_{ab}t_j \leq \frac{1}{\sqrt{2}}, \\ \leq s_j - \mu & \text{when } \Gamma_{ab}t_j \geq \sqrt{2}, \\ \approx s_j + \frac{e-1}{e+1} & \text{when } \Gamma_{ab}t_j \ll 1, \\ \approx s_j - \frac{1}{e+1} & \text{when } \Gamma_{ab}t_j \gg 1, \end{cases} \quad (\text{VII.3})$$

where  $\mu$  is a numerical constant,

$$\mu = 0.09. \quad (\text{VII.4})$$

Hence the random walk will tend to converge towards some integer  $s^*$  such that  $t^* = 2^{s^*}\tau_0$  satisfies  $\frac{1}{\sqrt{2}} \leq \Gamma_{ab}t^* \leq \sqrt{2}$ .

Note that the expected position of the random walk moves towards  $s^*$  at a rate that is lower-bounded by  $\mu$ . So, in order to go from  $s_0 = 0$  to  $s^* \approx -\log_2(\Gamma_{ab}\tau_0)$ , we expect that the random walk will take roughly  $\frac{1}{\mu}|s^*| = \frac{1}{\mu}|\log_2(\Gamma_{ab}\tau_0)|$  steps. It is easy to see that the stationary distribution of the random walk is centered around  $s^*$ , with exponentially decaying tails; hence, once the walk reaches  $s^*$ , it will remain concentrated around that point.

We now explain how to set the parameter  $N_{\text{steps}}$  so that, with high probability, after  $N_{\text{steps}}$  steps, the random walk will converge. Say we are given an upper bound  $h$  on the magnitude of  $s^*$ , i.e., we are promised that  $|s^*| \leq h$ , or equivalently, we are promised that  $2^{-h} \leq \Gamma_{ab}\tau_0 \leq 2^h$ . Then we will run the random walk for a number of steps

$$N_{\text{steps}} = \frac{h}{\mu} + \eta, \quad (\text{VII.5})$$

where  $\eta \geq 0$ . Here,  $\frac{h}{\mu}$  is (an upper bound on) the expected number of steps needed to reach  $s^*$ . We take an additional  $\eta$  steps to ensure that the walk does indeed reach  $s^*$  with high probability (we will show that the probability of failure decreases exponentially with  $\eta$ ).

We claim that, after  $N_{\text{steps}}$  steps, the final position of the walk is close to  $s^*$ , with exponentially decaying tail probabilities: for any  $\ell \geq 1$ ,

$$\begin{aligned} \Pr[|s_{N_{\text{steps}}} - s^*| \geq \ell | s_0 = 0] &\leq \frac{16}{\mu^2} \exp\left(-\frac{\mu(\mu+1)}{8}\ell + \frac{\mu^2}{4}\right) \\ &\quad + 2 \exp\left(-\frac{\mu}{16} \min\left\{\frac{\mu\eta}{h}, 1\right\}(\ell + \mu\eta)\right). \end{aligned} \quad (\text{VII.6})$$

In particular, when  $\eta \geq \frac{h}{\mu}$ , this bound can be slightly simplified:

$$\begin{aligned} \Pr[|s_{N_{\text{steps}}} - s^*| \geq \ell | s_0 = 0] &\leq \frac{16}{\mu^2} \exp\left(-\frac{\mu(\mu+1)}{8}\ell + \frac{\mu^2}{4}\right) + 2 \exp\left(-\frac{\mu}{16}(\ell + \mu\eta)\right). \end{aligned} \quad (\text{VII.7})$$



The bound (VII.6) is proved in Appendix D, using martingale techniques.

### B. Error Bound for $\hat{t}$

We now explain how to set the parameter  $N_{\text{trials}}$ , and derive an error bound on our estimator  $\hat{t}$ . The bound (VII.6) implies that the random variables  $\xi_r = s_{N_{\text{steps}}}$  are sub-exponential (see [27], section 5.2.4), and their sub-exponential norms are bounded by some constant  $\|\xi_r\|_{\psi_1} \leq K$ . Hence their average  $\xi$  satisfies a Bernstein-type concentration inequality (see [27], corollary 5.17): for every  $\delta \geq 0$ ,

$$\Pr[|\xi - s^*| \geq \delta] \leq 2 \exp(-c \min\{\frac{\delta^2}{K^2}, \frac{\delta}{K}\} N_{\text{trials}}), \quad (\text{VII.8})$$

where  $c > 0$  is a universal constant.

Now, for any  $\epsilon > 0$ , we set

$$N_{\text{trials}} = \frac{1}{\epsilon} \max\{\frac{K}{\delta}, \frac{K^2}{\delta^2}\} \log(\frac{2}{\epsilon}). \quad (\text{VII.9})$$

Then we have the following error bound on our estimator  $\hat{t} = 2^\xi \tau_0$ : with probability  $\geq 1 - \epsilon$ , we have  $|\xi - s^*| < \delta$ , which implies that

$$2^{-0.5-\delta} < \Gamma_{\text{ab}} \hat{t} < 2^{0.5+\delta}. \quad (\text{VII.10})$$

Assuming  $\delta < 1/2$ , this implies that  $\frac{1}{2} < \Gamma_{\text{ab}} \hat{t} < 2$ , as desired.

## VIII. EFFECT OF SPAM ERRORS

When the measurement protocols described in this paper are implemented in an experiment, errors may occur during state preparation and measurement (SPAM errors). We investigate the effect of these errors on estimating the decay rates  $\Gamma_{\text{ab}}$ . Let  $\rho_0$  and  $E_0$  denote the noiseless initial state and observable of interest, respectively. We have

$$\rho_0 = E_0 = \frac{1}{2}(|\mathbf{a}\rangle\langle\mathbf{a}| + |\mathbf{b}\rangle\langle\mathbf{b}| + |\mathbf{b}\rangle\langle\mathbf{a}| + |\mathbf{a}\rangle\langle\mathbf{b}|). \quad (\text{VIII.1})$$

We consider error channels  $\mathcal{E}_s$  and  $\mathcal{E}_m$  that act on state preparation and measurement operations as

$$\tilde{\rho} = \mathcal{E}_s(\rho_0) = \rho_0 + \delta\rho \quad (\text{VIII.2})$$

$$\tilde{E} = \mathcal{E}_m(E_0) = E_0 + \delta E, \quad (\text{VIII.3})$$

where  $\|\delta\rho\|_{\text{tr}} \leq \epsilon_s$  and  $\|\delta E\| \leq \epsilon_m$ , and  $\epsilon_s$  and  $\epsilon_m$  are small parameters. The outcome of the protocol is now given by

$$\tilde{P}_{\text{ab}}(t) = \text{Tr}[\tilde{E}\tilde{\mathcal{E}}_t(\tilde{\rho})]. \quad (\text{VIII.4})$$

where  $\mathcal{E}_t = \exp(\mathcal{L}t)$  is the evolution under the correlated dephasing noise (I.1).

We show that our protocol is robust against these kinds of errors, and for short times  $t$  the decay of  $\tilde{P}_{\text{ab}}(t)$  is still dominated by  $\Gamma_{\text{ab}}$ . Using Eqs. (VIII.2) and (VIII.3) we find that

$$\begin{aligned} \text{Tr}[\tilde{E}\tilde{\mathcal{E}}_t(\tilde{\rho})] &= \text{Tr}[E_0\mathcal{E}_t(\rho_0)] \\ &+ \text{Tr}[E_0\mathcal{E}_t(\delta\rho)] + \text{Tr}[\delta E\mathcal{E}_t(\rho_0)] \\ &+ \text{Tr}[\delta E\mathcal{E}_t(\delta\rho)] \end{aligned} \quad (\text{VIII.5})$$

The first term is the outcome without errors, and we have

$$\text{Tr}[E_0\mathcal{E}_t(\rho_0)] = \frac{1 + e^{-t\Gamma_{\text{ab}}}}{2}. \quad (\text{VIII.6})$$

We can find the effect of errors on the second and third terms, by considering the effect of  $\mathcal{E}_t$  on  $\rho_1$  and  $E_1$ . Specifically, we find

$$\text{Tr}[E_0\mathcal{E}_t(\delta\rho)] = \eta_s + \zeta_s e^{-\Gamma_{\text{ab}}t}, \quad (\text{VIII.7})$$

$$\text{Tr}[\delta E\mathcal{E}_t(\rho_0)] = \eta_m + \zeta_m e^{-\Gamma_{\text{ab}}t}, \quad (\text{VIII.8})$$

where  $\eta_{s,m}$  and  $\zeta_{s,m}$  are constants that are determined by  $\delta\rho$  and  $\delta E$ , for  $s$  and  $m$ , respectively. Therefore, these terms decay with the same rate as the first case and do not affect the exponential decay. However, the last term can, in principle, contain different decay rates and can cause deviation from a single exponential decay. We can bound the rate at which  $R(t) = \text{Tr}[\delta E\mathcal{E}_t(\delta\rho)]$  grows:

$$|\dot{R}(t)| = \left| \frac{\partial}{\partial t} \text{Tr}[\delta E\mathcal{E}_t(\delta\rho)] \right| \quad (\text{VIII.9})$$

$$= |\text{Tr}[\delta E\mathcal{L}(\delta\rho)]| \quad (\text{VIII.10})$$

$$\leq 2\epsilon_m\epsilon_s(n+s), \quad (\text{VIII.11})$$

see Appendix E for the proof. Therefore, we find

$$\tilde{P}_{\text{ab}}(t) = \frac{1}{2}(1 + \eta_s + \eta_m + (1 + \zeta_s + \zeta_m)e^{-\Gamma_{\text{ab}}t}) + R(t), \quad (\text{VIII.12})$$

The deviations from a single exponential decay are attributed to  $R(t)$ . Using Eqs. (VIII.11) and (VIII.12) we can see that the decay rate of  $\tilde{P}_{\text{ab}}(t)$  is dominated by  $\Gamma_{\text{ab}}$  for evolution times  $t \lesssim 1/(2\epsilon_s\epsilon_m(n+s))^{-1}$ .

We also use numerical simulations to investigate the effect of SPAM errors on estimates of the decay rate  $\Gamma_{\text{ab}}$ . We simulate SPAM errors by applying random error channels  $\mathcal{E}_s$  and  $\mathcal{E}_m$ , whose strengths are controlled by a parameter  $\Delta$  (see Appendix B for details). We then compare the decay of  $\tilde{P}_{\text{ab}}(t)$  with and without SPAM errors, for different values

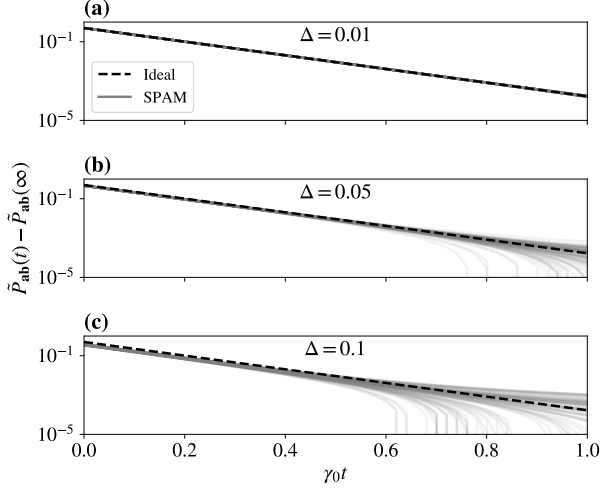


FIG. VIII.1. Exponential decay of  $\tilde{P}_{\mathbf{ab}}(t)$  when there are SPAM errors. We simulate the decay of a 3-qubit GHZ state. We choose a correlation matrix  $C$  with uniform single-qubit dephasing rate ( $c_{ii} = \gamma_0$ ) and nearest-neighbor correlations  $c_{i,i\pm 1} = \gamma_0/4$ . The panels show the decay of  $\tilde{P}_{\mathbf{ab}}(t)$  under different noise strengths: (a)  $\Delta = 0.01$ , (b)  $\Delta = 0.05$ , (c)  $\Delta = 0.1$ . (See Appendix B for the precise definition of  $\Delta$ .) The dashed lines show the decay with no SPAM errors, and solid lines show the decay with SPAM errors from randomly-sampled error channels. The solid lines resemble the dashed lines for short evolution times  $t$ .

of  $\Delta$ . We observe that, for short times  $t$ , the decay with SPAM errors matches the decay without SPAM errors, i.e., the decay rate is dominated by  $\Gamma_{\mathbf{ab}}$ , see Fig. VIII.1. This is consistent with our theoretical analysis.

## IX. GENERALIZATIONS

We now sketch how our method for learning sparse correlated dephasing noise can be extended to the most general case of the master equation (I.1), where the matrix  $C$  is complex, and there is an additional Hamiltonian term  $H_s$ .

### A. Complex Decay Rates

The complete dynamics imposed by the environment on the system can have a coherent evolution in addition to the decay. The evolution of the system  $\frac{d\rho}{dt} = \mathcal{L}(\rho)$  is in general given by the Lindblad

generator

$$\mathcal{L}(\rho) = -i[\rho, H_s] + \sum_{l,m} c_{lm} \left( Z_l \rho Z_m - \frac{1}{2} \{Z_l Z_m, \rho\} \right), \quad (\text{IX.1})$$

where

$$H_s = \sum_{l,m} r_{lm} Z_l Z_m \quad (\text{IX.2})$$

is sometimes called the (generalized) Lamb shift Hamiltonian, and  $C = (c_{lm})$  is now a complex matrix [11] (see Appendix A). We can decompose  $C$  as

$$C = V + iT, \quad (\text{IX.3})$$

where the real and imaginary parts are separated into  $V = (v_{lm})$ , a real symmetric matrix, and  $T = (t_{lm})$ , a real skew-symmetric matrix, respectively. Moreover, the Lamb shift Hamiltonian can be encoded in the symmetric matrix  $R = (r_{lm})$ .

We now show how the operator  $\mathcal{L}$  acts on the off-diagonal matrix elements  $|\mathbf{a}\rangle\langle\mathbf{b}|$ . This involves “decay rates” that depend on  $R$ ,  $V$  and  $T$  in a simple way, although these “decay rates” now complex rather than real. Specifically,

$$\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = (-\Gamma_{\mathbf{ab}} + i\Omega_{\mathbf{ab}})|\mathbf{a}\rangle\langle\mathbf{b}|, \quad (\text{IX.4})$$

where  $\Gamma_{\mathbf{ab}}$  and  $\Omega_{\mathbf{ab}}$  are real numbers that capture the decay and oscillations of the matrix element, respectively. We follow the convention defined in Section II A for states  $|\mathbf{a}\rangle$  and  $|\mathbf{b}\rangle$ . We remind the reader that  $Z_i|\mathbf{a}\rangle = \alpha_i|\mathbf{a}\rangle$ , and similarly for  $|\mathbf{b}\rangle$  and  $\beta_i$ . Therefore, separating the real and imaginary part of  $\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|)$  we find that

$$\begin{aligned} \Gamma_{\mathbf{ab}} &= -\sum_{l,m} v_{lm} (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) \\ &= \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\beta})^T V (\boldsymbol{\alpha} - \boldsymbol{\beta}), \end{aligned} \quad (\text{IX.5})$$

where we used the fact that  $V$  is symmetric. Similarly, we find

$$\begin{aligned} \Omega_{\mathbf{ab}} &= -\sum_{l,m} r_{lm} (\alpha_l \alpha_m - \beta_l \beta_m) \\ &\quad + \sum_{l,m} v_{lm} (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) \\ &= -(\boldsymbol{\alpha}^T R \boldsymbol{\alpha} - \boldsymbol{\beta}^T R \boldsymbol{\beta}) + \frac{1}{2} (\boldsymbol{\alpha}^T T \boldsymbol{\beta} - \boldsymbol{\beta}^T T \boldsymbol{\alpha}), \end{aligned} \quad (\text{IX.6})$$

where we used the fact that  $R$  and  $T$ , are symmetric and skew-symmetric matrices, respectively. (See Appendix A 2 for details.)

Therefore, the coherences, that is, the  $|\mathbf{a}\rangle\langle\mathbf{b}|$  elements of the density matrix  $\rho(t)$ , exhibit both oscillations (at a frequency  $\Omega_{\mathbf{ab}}$ ) and exponential decay (at a rate  $\Gamma_{\mathbf{ab}}$ ) as a function of  $t$ . Note that it is possible to measure both  $\Gamma_{\mathbf{ab}}$  and  $\Omega_{\mathbf{ab}}$ , by estimating the matrix elements of  $\rho(t)$  at different times  $t$ , using the same types of Ramsey experiments outlined in Section II. (In particular, one can extract these parameters using standard techniques for analyzing spectral lines in atomic physics. Here, the squared magnitude of the Fourier transform of the measurement time series is a Lorentzian function, the center of the Lorentzian peak gives the oscillation frequency, and the width of the peak gives the decay rate [50].)

Given estimates of  $\Gamma_{\mathbf{ab}}$  and  $\Omega_{\mathbf{ab}}$ , we can extend our compressed sensing method to recover both the correlation matrix  $C$  and the Hamiltonian  $H_s$ . As before, we use  $m \sim s \log n$  or  $m \sim s \log^4 n$  measurement settings; for each measurement setting, we choose  $\mathbf{a}$  and  $\mathbf{b}$  uniformly at random in  $\{0, 1\}^n$ . From estimates of  $\Gamma_{\mathbf{ab}}$ , we can recover  $V$ , the real part of  $C$ , exactly as before (this follows from equation (IX.5)). In a similar manner, given estimates of  $\Omega_{\mathbf{ab}}$ , we want to recover  $T$ , the imaginary part of  $C$ , as well as  $R$ , the matrix that encodes the Hamiltonian  $H_s$ . This is possible, because equation (IX.6) represents a measurement of  $R$  and  $T$  that has the needed isotropy and incoherence properties, as we will show below.

### B. Isotropy and Incoherence of the Measurements $\Omega_{\mathbf{ab}}$

We consider  $\Omega_{\mathbf{ab}}$ , viewed as a measurement of  $R$  and  $T$  in equation (IX.6), with  $\mathbf{a}$  and  $\mathbf{b}$  chosen uniformly at random in  $\{0, 1\}^n$ . We show that this random measurement has the same isotropy and incoherence properties as before, and hence our compressed sensing method will still succeed using these measurements. The incoherence property is easy to see, but some work is required to show the isotropy property.

The measurement  $\Omega_{\mathbf{ab}}$  acts on  $T$  and  $R$  as

$$\Omega_{\mathbf{ab}} = \begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{\beta}^T \end{bmatrix} \begin{bmatrix} -R & \frac{1}{2}T \\ -\frac{1}{2}T & R \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}. \quad (\text{IX.7})$$

Similar to the analysis in Section VD, it is advantageous to consider the effect of measurements if we enforce the symmetries of  $R$  and  $T$ . Note that  $T$  and  $R$  are real skew-symmetric and symmetric matrices,

respectively. Moreover, they are both traceless. Therefore, we have

$$\Omega_{\mathbf{ab}} = \sum_{i < j} (\alpha_i \beta_j - \beta_i \alpha_j) T_{ij} + 2 \sum_{i < j} (\alpha_i \alpha_j - \beta_i \beta_j) R_{ij} \quad (\text{IX.8})$$

As before, let  $\text{uvec}$  be the linear operator that returns the upper-triangular part of a matrix (not including the diagonal elements), that is,

$$\text{uvec} : R \mapsto (R_{ij})_{i < j}. \quad (\text{IX.9})$$

Then  $\Omega_{\mathbf{ab}}$  can be expressed as

$$\Omega_{\mathbf{ab}} = \mathbf{q} \begin{bmatrix} \text{uvec}(T) \\ 2 \text{uvec}(R) \end{bmatrix}, \quad (\text{IX.10})$$

similar to Eq. (V.26), where  $\mathbf{q}$  is a row vector of the form

$$\mathbf{q} = \left[ \text{uvec}(\boldsymbol{\alpha}\boldsymbol{\beta}^T - \boldsymbol{\beta}\boldsymbol{\alpha}^T), \text{uvec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \boldsymbol{\beta}\boldsymbol{\beta}^T) \right]. \quad (\text{IX.11})$$

Note that as described in Section III A,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are chosen uniformly and independently at random from  $\{1, -1\}^n$ . It is then straightforward to see that  $\mathbf{q}$  in this case satisfies the incoherence property (V.30) with  $\mu = 1$ . Furthermore, one can check that  $\mathbf{q}$  is centered and isotropic (up to a normalization factor of  $\sqrt{2}$ ), since:

$$\begin{cases} \mathbb{E}[\alpha_i \alpha_j] = \mathbb{E}[\alpha_i \beta_j] = \mathbb{E}[\beta_i \beta_j] = 0 \\ \mathbb{E}[(\alpha_i \beta_j - \beta_i \alpha_j)(\alpha_k \beta_l - \beta_k \alpha_l)] = 2\delta_{ik}\delta_{jl} \\ \mathbb{E}[(\alpha_i \alpha_j - \beta_i \beta_j)(\alpha_k \alpha_l - \beta_k \beta_l)] = 2\delta_{ik}\delta_{jl} \\ \mathbb{E}[(\alpha_i \beta_j - \beta_i \alpha_j)(\alpha_k \alpha_l - \beta_k \beta_l)] = 0 \end{cases}, \quad (\text{IX.12})$$

where it is assumed that  $i < j$  and  $k < l$  in all cases. The second and the third lines in the above equation capture the correlations in the measurements of  $T$  and  $R$ , respectively, and the last line captures the cross-correlations between the two measurements (see Appendix F for more details).

## X. OUTLOOK

In this paper we have demonstrated a new paradigm for characterizing spatially-correlated noise in many-body quantum systems, where one assumes that the correlations are sparse, and one uses compressed sensing to characterize them efficiently. Our method applies to correlated dephasing noise, and is capable of detecting long-range correlations,

with provable recovery guarantees, even in the presence of SPAM errors.

There are several possible generalizations of our method. First, we expect that our method can be applied to a larger class of Markovian noise processes, whose Lindblad operators are Hermitian and commuting. It may also be possible to combine our methods with multi-qubit noise spectroscopy [51], in order to estimate spatial correlations in non-Markovian noise processes. Finally, it is an interesting question whether our methods can be combined with the more combinatorial techniques of [10] for estimating sparse Pauli channels.

Our work touches on several other broad questions. First, we expect our protocol to be useful for characterizing dephasing in the presence of weak relaxation processes, i.e., when the relaxation time is much longer than the dephasing time. However, the harder case of concurrent relaxation and dephasing remains open, and is a subject of future research.

Also, while we use entangled states to probe the correlations in the noise process, it is intriguing to ask whether it is possible to achieve a similar scaling using non-entangled states. This motivates the more general question of whether entanglement can help noise spectroscopy in the same way that it is helpful in metrology [52].

## ACKNOWLEDGEMENTS

We thank Paola Cappellaro, Alexey Gorshkov, Amir Kalev, David Layden, Ingo Roth, and Lorenza Viola for helpful discussions. We thank Victor Albert, Richard Kueng and Zachary Levine for helpful comments on an earlier draft of this paper.

This research was supported in part by the AFOSR MURI project ‘‘Scalable Certification of Quantum Computing Devices and Networks,’’ ARO-MURI, and Northrop Grumman. This work was completed while AS was at the University of Maryland, before moving to the University of Chicago.

## Appendix A: Sparse Correlated Dephasing

### 1. Microscopic Derivation

We consider a model of spatially correlated noise, which is described by the following master equation:

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = i[\rho, H_s] + \sum_{j,k=1}^n c_{jk} \left( s_k \rho s_j^\dagger - \frac{1}{2} \{s_j^\dagger s_k, \rho\} \right). \quad (\text{A.1})$$

Here the system consists of  $n$  qubits,  $H_s$  is the system Hamiltonian, and  $s_j$  and  $s_k$  are operators that act on the  $j$ 'th and  $k$ 'th qubits, respectively. We consider the case of dephasing noise, where  $H_s = 0$ , and  $s_j = Z_j$  (the Pauli  $\sigma_z$  operator acting on the  $j$ 'th qubit). The noise is then fully described by the matrix of coupling coefficients  $C = (c_{jk}) \in \mathbb{C}^{n \times n}$ . In order for the time evolution of the system to be a completely positive map, the matrix  $C$  must be positive semidefinite [21].

We will be interested in the case where the matrix  $C$  is *sparse*, in the sense that most of the off-diagonal elements are 0. This can arise when the qubits are coupled to a bath, in such a way that most of the time, different qubits are coupled to *different* degrees of freedom in the bath, but occasionally it happens that two or more qubits couple to the *same* degree of freedom in the bath.

To motivate the above intuition, we follow Ref. [14] and show how the master equation (A.1) can be derived from a commonly-occurring microscopic model. Here, multiple qubits are coupled to a bosonic bath through the interaction Hamiltonian

$$H_I(t) = \sum_{l,\mathbf{k}} Z_l \left( g_{\mathbf{k}}^l b_{\mathbf{k}}^\dagger(t) + g_{\mathbf{k}}^{*l} b_{\mathbf{k}}(t) \right), \quad (\text{A.2})$$

where  $b_{\mathbf{k}}(t)$  are bath operators. If the bath is quantum mechanical, then  $b_{\mathbf{k}}(t) = e^{-i\omega_{\mathbf{k}}t} b_{\mathbf{k}}$ , where  $b_{\mathbf{k}}$  is a bosonic annihilation operator corresponding to the bath mode  $\mathbf{k}$ , with frequency  $\omega_{\mathbf{k}}$ . If the bath is classical, then  $b_{\mathbf{k}}(t)$  is a stochastic variable corresponding to the fluctuations of the external magnetic field [51].

The time evolution operator is then given by [14]

$$U_I(t) = \exp[i\Phi_{\omega_{\mathbf{k}}}(t)] \exp\left[ \sum_{l,\mathbf{k}} Z_l \{ \xi_{\mathbf{k}}^l(t) b_{\mathbf{k}}^\dagger - \xi_{\mathbf{k}}^{*l}(t) b_{\mathbf{k}} \} \right] \quad (\text{A.3})$$

where

$$\Phi_{\omega_{\mathbf{k}}}(t) = \sum_{l,\mathbf{k}} |Z_l g_{\mathbf{k}}^l|^2 \frac{\omega_{\mathbf{k}} t - \sin(\omega_{\mathbf{k}} t)}{\omega_{\mathbf{k}}^2}, \quad (\text{A.4})$$

$$\xi_{\mathbf{k}}^l = g_{\mathbf{k}}^l \frac{1 - e^{i\omega_{\mathbf{k}} t}}{\omega_{\mathbf{k}}}. \quad (\text{A.5})$$

The phase term  $\Phi_{\omega_{\mathbf{k}}}(t)$  only appears when the bath is quantum mechanical, and is the result of the non-commutative nature of the bath operators. This term is also known as the Lamb shift. We assume that if qubit  $l$  at position  $\mathbf{x}_l$  is coupled to the bath mode  $k$ , then the coupling is of the form  $g_{\mathbf{k}}^l = g_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}_l}$ , and if the qubit is not coupled to that mode, then  $g_{\mathbf{k}}^l = 0$ .

Defining  $|\mathbf{a}\rangle = |a_1, a_2, \dots, a_n\rangle$ ,  $|\mathbf{b}\rangle = |b_1, b_2, \dots, b_n\rangle$ , with  $a_m, b_m \in \{0, 1\}$ , it follows that

$$\langle \mathbf{a} | \rho(t) | \mathbf{b} \rangle = \exp\left[-\frac{1}{2} \sum_{m,l} \tilde{v}_{m,l}(t) (\alpha_m - \beta_m) (\alpha_l - \beta_l)\right] \quad (\text{A.6})$$

$$\begin{aligned} & \times \exp\left[i \sum_{m,l} \tilde{t}_{m,l}(t) (\alpha_m \beta_l)\right] \\ & \times \exp\left[-i \sum_{m,l} \tilde{r}_{m,l}(t) (\alpha_m \alpha_l - \beta_m \beta_l)\right] \\ & \times \langle \mathbf{a} | \rho(0) | \mathbf{b} \rangle, \end{aligned} \quad (\text{A.7})$$

where we defined  $\alpha_m = (-1)^{a_m}$ ,  $\beta_m = (-1)^{b_m}$ , and

$$\tilde{v}_{l,m}(t) = 2 \sum_{\mathbf{k}} |g_{\mathbf{k}}|^2 \frac{1 - \cos(\omega_{\mathbf{k}} t)}{\omega_{\mathbf{k}}^2} \coth\left(\frac{\beta \omega_{\mathbf{k}}}{2}\right) \cos(\mathbf{k} \cdot \mathbf{x}_{l,m}) \quad (\text{A.8})$$

$$\tilde{t}_{l,m}(t) = -2 \sum_{\mathbf{k}} |g_{\mathbf{k}}|^2 \frac{1 - \cos(\omega_{\mathbf{k}} t)}{\omega_{\mathbf{k}}^2} \sin(\mathbf{k} \cdot \mathbf{x}_{l,m}) \quad (\text{A.9})$$

$$\tilde{r}_{l,m}(t) = - \sum_{\mathbf{k}} |g_{\mathbf{k}}|^2 \frac{\omega_{\mathbf{k}} t - \sin(\omega_{\mathbf{k}} t)}{\omega_{\mathbf{k}}^2} \cos(\mathbf{k} \cdot \mathbf{x}_{l,m}), \quad (\text{A.10})$$

where  $\mathbf{x}_{l,m} = \mathbf{x}_l - \mathbf{x}_m$ . Note that  $\tilde{t}_{l,m}(v) = \tilde{v}_{m,l}(t)$ ,  $\tilde{t}_{l,m}(t) = -\tilde{t}_{m,l}(t)$ , and  $\tilde{r}_{l,m}(t) = \tilde{r}_{m,l}(t)$ . Therefore, it is illustrative to express Eq. (A.6) as [53]

$$\begin{aligned} \frac{d\rho}{dt} &= i[\rho, \sum_{l,m} r_{l,m}(t) Z_l Z_m] \\ &+ \sum_{l,m} c_{l,m}(t) \left( Z_l \rho Z_m - \frac{1}{2} \{Z_l Z_m, \rho\} \right), \end{aligned} \quad (\text{A.11})$$

where we defined  $\tilde{c}_{l,m}(t) = \frac{1}{2} \tilde{v}_{l,m}(t) - i \tilde{t}_{l,m}(t)$  and  $\alpha_{l,m}(t) = \frac{\partial}{\partial t} \tilde{c}_{l,m}(t)$  for  $o \in \{v, t, r, c\}$ . We can see that Eq. (A.11), is in the Lindblad form of Eq. (I.1), but with time dependent coefficients.

It should be noted that  $v_{l,m}(t)$ ,  $t_{l,m}(t)$ , and  $r_{l,m}(t)$  vanish if qubits  $m$  and  $n$  do not share a bath mode, that is, if there does not exist any  $\mathbf{k}$  such that  $g_{\mathbf{k}}^l$  and  $g_{\mathbf{k}}^m$  are both non-zero. Therefore, in cases where only a few qubit pairs have shared bath modes, these coefficients are *sparse*.

A special case that is extensively studied in the literature [12–14] is collective decoherence, where the distance between the qubits is much smaller than the wavelength of bath modes,  $\exp(i\mathbf{k} \cdot \mathbf{x}_{l,m}) \approx 1$ , and consequently  $v_{l,m} \approx 0$ . In this limit, qubits that are coupled to the same modes are maximally correlated.

Moreover, in the high-temperature limit, where the temperature of the environment is much larger than the energy of the highest-frequency bath modes, the decaying part of the evolution due to  $\tilde{t}_{l,m}$  dominate the coherent evolution that comes from the Lamb shift term  $\tilde{r}_{l,m}$  [14]. In that limit, the decoherence dynamics is solely determined by the *real* coefficients  $c_{l,m}$ . This is similar to the case of a classical bath (with scalar stochastic  $b_{\mathbf{k}}$ 's), where again the dynamics solely depend on *real* coefficients  $c_{l,m}$  without the Lamb shift Hamiltonian.

Finally, we note that within Born-Markov approximation [11], the coefficients  $t_{l,m}$ ,  $v_{l,m}$ , and  $r_{l,m}$  are time-independent and we recover Eq. I.1 with the Lamb shift Hamiltonian  $H_s = \sum_{l,m} r_{l,m} Z_l Z_m$  [11]. The validity of this approximation, depends on the details of the bath and the time scales under which the system is being studied.

In this work we assume that the Markovian approximation is valid and the master equation has a time-independent generator  $\mathcal{L}$ . In this regime, Eq. (A.11) coincides with Eq. (IX.1).

## 2. Derivation of the Dynamics

In this section we present the full details of the derivation of the dynamics under the master equation (IX.1). Note that  $\mathcal{L}$  does not map  $|\mathbf{a}\rangle\langle\mathbf{b}|$  to another  $|\mathbf{a}'\rangle\langle\mathbf{b}'|$ . Additionally,  $\mathcal{L}$  is linear. Therefore we can treat the evolution independently for each element  $|\mathbf{a}\rangle\langle\mathbf{b}|$  of the density matrix  $\rho$ , and obtain  $\rho(t) = \sum_{\mathbf{a},\mathbf{b}} \langle \mathbf{a} | \rho(0) | \mathbf{b} \rangle \exp[(i\Omega_{\mathbf{ab}} t - \Gamma_{\mathbf{ab}} t] | \mathbf{a} \rangle \langle \mathbf{b} |$ .

$$\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = -i[|\mathbf{a}\rangle\langle\mathbf{b}|, \sum_{l,m} r_{lm} Z_l Z_m] + \sum_{l,m} c_{lm} (Z_l |\mathbf{a}\rangle\langle\mathbf{b}| Z_m - \frac{1}{2} \{Z_l Z_m, |\mathbf{a}\rangle\langle\mathbf{b}|\}) \quad (\text{A.12})$$

$$= \left[ -i \sum_{l,m} r_{lm} (\alpha_l \alpha_m - \beta_l \beta_m) + \sum_{l,m} c_{lm} (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) \right] |\mathbf{a}\rangle\langle\mathbf{b}|. \quad (\text{A.13})$$

We now use the symmetries of  $C$  and express the second sum as

$$\sum_{l,m} c_{lm} (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) = \sum_{l,m} (v_{lm} + it_{lm}) (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) \quad (\text{A.14})$$

$$= \sum_{l,m} \frac{1}{2} (v_{lm} + it_{lm}) (\alpha_l \beta_m - \frac{1}{2} \alpha_l \alpha_m - \frac{1}{2} \beta_l \beta_m) + \sum_{l,m} \frac{1}{2} (v_{lm} - it_{lm}) (\alpha_m \beta_l - \frac{1}{2} \alpha_m \alpha_l - \frac{1}{2} \beta_m \beta_l) \quad (\text{A.15})$$

$$= \frac{v_{lm}}{2} (\alpha_l \beta_m + \alpha_m \beta_l - \alpha_l \alpha_m - \beta_l \beta_m) + \frac{it_{lm}}{2} (\alpha_l \beta_m - \alpha_m \beta_l) \quad (\text{A.16})$$

$$= -\frac{1}{2} v_{lm} (\alpha_l - \beta_l) (\alpha_m - \beta_m) + \frac{i}{2} t_{lm} (\alpha_l \beta_m - \alpha_m \beta_l) \quad (\text{A.17})$$

We express the above results compactly as  $\mathcal{L}(|\mathbf{a}\rangle\langle\mathbf{b}|) = (-\Gamma_{\mathbf{ab}} + i\Omega_{\mathbf{ab}})|\mathbf{a}\rangle\langle\mathbf{b}|$ , where

$$\Gamma_{\mathbf{ab}} = \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\beta})^T V (\boldsymbol{\alpha} - \boldsymbol{\beta}), \quad (\text{A.18})$$

$$\Omega_{\mathbf{ab}} = \frac{1}{2} (\boldsymbol{\alpha}^T T \boldsymbol{\beta} - \boldsymbol{\beta}^T T \boldsymbol{\alpha}) - (\boldsymbol{\alpha}^T R \boldsymbol{\alpha} - \boldsymbol{\beta}^T R \boldsymbol{\beta}). \quad (\text{A.19})$$

## Appendix B: Details of Numerical Simulations

### 1. Compressed Sensing of Correlated Dephasing Noise

Here we provide some additional details about the numerical simulations in Section IV.

Assume that we have  $n$  qubits, whose individual dephasing rates is fully characterized. However, correlations in noise cannot be observed in single qubit measurements. We assume that there are  $s'$  pairs of qubits that are correlated. To generate a correlation matrix  $C$  that is positive semidefinite with a controllable number of non-zero off-diagonal elements  $s = 2s'$ , we choose

$$\begin{cases} c_{ii} = 2 & i = 1, \dots, n \\ c_{ij} = c_{ji} = \frac{1}{2} & 1 \leq i \leq s' \text{ and } j = i + 1. \end{cases} \quad (\text{B.1})$$

We then remove the spatial structure in the matrix by randomly permuting the rows and columns of  $C$ , that is we map  $c_{ij}$  to  $c_{\pi(i),\pi(j)}$  with  $\pi \in S_n$ . This procedure ensures that the eigenvalues of  $C$  are non-negative.

We then generate noiseless measurements by choosing  $m$  samples of  $\mathbf{a}$  and  $\mathbf{b}$ . Finally, we assume that the diagonal elements of  $C$  are known, and we solve the convex optimization problem (III.5)-(III.7) using CVXPY, a convex optimization package for Python [47, 48]. The scaling of the reconstruction error  $\|W^{(opt)} - C\|_\infty$  as a function of  $m$  for different values of  $n$  and  $s$  is illustrated in Fig. I.2 (a-b).

We also investigate the effect of noisy measurements on the recovery error. We add a random vector  $\mathbf{e}$ , whose entries are independent Gaussian random variables with mean 0 and standard deviation  $\sigma_\epsilon$ , to the measurement vector  $\mathbf{h}$ . Therefore we replace the constraint (III.7) with (III.12). Additionally, we assume that the diagonals of  $C$  are known and hence  $\epsilon_1 = 0$  in (III.11). We choose  $\epsilon_2 = \sqrt{m} \sigma_\epsilon$  and solve the convex program. The scaling of the reconstruction error  $\|W^{(opt)} - C\|_\infty$  as a function of  $\sigma_\epsilon$  is shown in Fig. I.2 (c).

## 2. State Preparation and Measurement Errors

We provide some additional details about the simulations in Section VIII, showing the effects of SPAM errors. We generate error channels  $\mathcal{E}_s$  and  $\mathcal{E}_m$  that act on  $n$  qubits, by adding  $n$  auxiliary qubits in the state  $|0\rangle^{\otimes n}$ , applying unitary operations  $U_s$  and  $U_m$  on all  $2n$  qubits, and tracing out the  $n$  auxiliary qubits. We control the strength of these error channels by a parameter  $\Delta$ , and we set  $U_s = \exp(i\Delta H)$  where  $H_s$  is a Hermitian operator on  $2n$  qubits. To generate different error channels at random, we choose  $H_s = \frac{1}{2}(M_s + M_s^\dagger)$ , where elements of  $M_s$  are complex numbers whose real and imaginary parts are chosen uniformly at random from  $[0, 1]$ . We construct  $U_m$  in a similar way.

### Appendix C: Setting the Regularization Parameter $\lambda$

Here we show that, if  $\lambda$  is set according to equation (V.20), then it satisfies equations (V.15) and (V.16).

In order to prove these bounds, we will want to use Hoeffding-type large deviation bounds on  $\mathbf{u}$  and  $\mathbf{v}$ . To this end, we first need to upper-bound the quantities  $\max_{1 \leq j \leq n} \|u_j\|_{\psi_2}$  and  $\max_{1 \leq j \leq m} \|v_j\|_{\psi_2}$ . We can do this as follows. First write

$$\max_{1 \leq j \leq n} \|u_j\|_{\psi_2} \leq \delta_1 \|\text{diag}(C)\|_\infty, \quad (\text{C.1})$$

and

$$\|\text{diag}(C)\|_\infty \leq \|\mathbf{g}\|_\infty + \|\mathbf{u}\|_\infty. \quad (\text{C.2})$$

We can upper-bound  $\|\mathbf{u}\|_\infty$  by using the union bound over all  $j \in \{1, \dots, n\}$ , and bounding each coordinate  $u_j$  using the definition of the sub-Gaussian norm (see Definition 5.7 in [27]), as follows. Let  $c_0 > 0$  be the universal constant in Definition 5.7 in [27]. For any  $t \geq 0$ , we have

$$\begin{aligned} & \Pr[\|\mathbf{u}\|_\infty > (t+1)\sqrt{\ln(n)/c_0} \delta_1 \|\text{diag}(C)\|_\infty] \\ & \leq \sum_{j=1}^n \Pr[|u_j| > (t+1)\sqrt{\ln(n)/c_0} \delta_1 \|\text{diag}(C)\|_\infty] \\ & \leq \sum_{j=1}^n \Pr[|u_j| > (t+1)\sqrt{\ln(n)/c_0} \delta_1 c_{jj}] \\ & \leq n \cdot e \cdot \exp(-(t+1)^2 \ln(n)) \\ & = e \cdot n^{-t(t+2)}. \end{aligned} \quad (\text{C.3})$$

Now let  $t = 1$ . Then we have that

$$\|\mathbf{u}\|_\infty \leq \epsilon'_1 \|\text{diag}(C)\|_\infty, \quad \epsilon'_1 := 2\sqrt{\ln(n)/c_0} \delta_1, \quad (\text{C.4})$$

with failure probability at most  $e/n^3$ . Combining this with (C.2) and (C.1), we get:

$$\|\text{diag}(C)\|_\infty \leq \frac{1}{1-\epsilon'_1} \|\mathbf{g}\|_\infty, \quad (\text{C.5})$$

provided that  $\epsilon'_1 < 1$ , and

$$\max_{1 \leq j \leq n} \|u_j\|_{\psi_2} \leq \frac{\delta_1}{1-\epsilon'_1} \|\mathbf{g}\|_\infty =: \epsilon''_1, \quad (\text{C.6})$$

which is our desired result. We can use a similar argument to bound the  $v_j$ , which yields:

$$\|\mathbf{v}\|_\infty \leq \epsilon'_2 \|\Phi(C)\|_\infty, \quad \epsilon'_2 := 2\sqrt{\ln(m)/c_0} \delta_2, \quad (\text{C.7})$$

with failure probability at most  $e/m^3$ ,

$$\|\Phi(C)\|_\infty \leq \frac{1}{1-\epsilon'_2} \|\mathbf{h}\|_\infty, \quad (\text{C.8})$$

provided that  $\epsilon'_2 < 1$ , and

$$\max_{1 \leq j \leq m} \|v_j\|_{\psi_2} \leq \frac{\delta_2}{1-\epsilon'_2} \|\mathbf{h}\|_\infty =: \epsilon''_2. \quad (\text{C.9})$$

We are now ready to prove (V.15). We will bound  $\|\text{uvec}(\Phi_D^\dagger \mathbf{z})\|_\infty$ , by splitting it into two terms, using (V.17), as follows:

$$\begin{aligned} \|\text{uvec}(\Phi_D^\dagger \mathbf{z})\|_\infty & \leq \|\text{uvec}(\Phi_D^\dagger \mathbf{v})\|_\infty \\ & \quad + \|\text{uvec}(\Phi_D^\dagger \Phi(\text{diag}(\mathbf{u}))\|_\infty. \end{aligned} \quad (\text{C.10})$$

We can bound the first term in (C.10) as follows:

$$\begin{aligned} \|\text{uvec}(\Phi_D^\dagger \mathbf{v})\|_\infty & = \max_{1 \leq j < j' \leq n} |\mathbf{e}_j^T (\Phi_D^\dagger \mathbf{v}) \mathbf{e}_{j'}| \\ & = \max_{j < j'} |\mathbf{v}^T \Phi_D (\mathbf{e}_j \mathbf{e}_{j'}^T)| \\ & = \max_{j < j'} |\mathbf{v}^T \xi^{(j,j')}|, \end{aligned} \quad (\text{C.11})$$

where we defined

$$\xi^{(j,j')} = \Phi_D (\mathbf{e}_j \mathbf{e}_{j'}^T) \in \mathbb{R}^m, \quad 1 \leq j < j' \leq n, \quad (\text{C.12})$$

where  $\mathbf{e}_j$  denotes the  $j$ 'th standard basis vector in  $\mathbb{R}^n$ . Note that the  $k$ 'th entry of  $\xi^{(j,j')}$  is given by

$$\xi_k^{(j,j')} = 4(\mathbf{r}^{(k)})^T \mathbf{e}_j \mathbf{e}_{j'}^T \mathbf{r}^{(k)} = 4r_j^{(k)} r_{j'}^{(k)}, \quad (\text{C.13})$$

hence  $\xi^{(j,j')}$  is bounded by

$$\|\xi^{(j,j')}\|_2 \leq 4\sqrt{m}. \quad (\text{C.14})$$

We can then upper-bound  $\|\text{uvec}(\Phi_D^\dagger \mathbf{v})\|_\infty$ , by using a union bound over all  $1 \leq j < j' \leq n$ , and a Hoeffding-type large deviation bound (Proposition 5.10 in [27]). Let  $c' > 0$  be the universal constant in Proposition 5.10 in [27]. Then for any  $t \geq 0$ , we have:

$$\begin{aligned} \Pr[\|\text{uvec}(\Phi_D^\dagger \mathbf{v})\|_\infty > 4\sqrt{m}(t + \sqrt{2})\epsilon_2''' \sqrt{\ln(n)/c'}] \\ \leq \sum_{j < j'} \Pr[|\mathbf{v}^T \xi^{(j,j')}| > 4\sqrt{m}(t + \sqrt{2})\epsilon_2''' \sqrt{\ln(n)/c'}] \\ \leq \frac{1}{2}n(n-1) \cdot e \cdot \exp(-(t + \sqrt{2})^2 \ln n) \\ \leq \frac{1}{2}e \cdot n^{-t(t+2\sqrt{2})}. \end{aligned} \quad (\text{C.15})$$

Now let  $t = 1$ . Then we have

$$\|\text{uvec}(\Phi_D^\dagger \mathbf{v})\|_\infty \leq 4\sqrt{m}(1 + \sqrt{2})\epsilon_2''' \sqrt{\ln(n)/c'}, \quad (\text{C.16})$$

with failure probability at most  $\frac{1}{2}e \cdot n^{-(1+2\sqrt{2})}$ .

We now bound the second term in (C.10), using a similar approach:

$$\begin{aligned} \|\text{uvec}(\Phi_D^\dagger \Phi(\text{diag}(\mathbf{u})))\|_\infty \\ = \max_{1 \leq j < j' \leq n} |\mathbf{e}_j^T (\Phi_D^\dagger \Phi(\text{diag}(\mathbf{u}))) \mathbf{e}_{j'}| \\ = \max_{j < j'} |\mathbf{u}^T \Phi_G^\dagger \Phi_D(\mathbf{e}_j \mathbf{e}_{j'}^T)| \\ = \max_{j < j'} |\mathbf{u}^T \Phi_G^\dagger \xi^{(j,j')}|, \end{aligned} \quad (\text{C.17})$$

where  $\Phi_G$  is the submatrix of  $\Phi$  that contains those columns of  $\Phi$  that are indexed by the subset  $G = \{(j, j) \mid j = 1, \dots, n\}$ . We can bound  $\Phi_G^\dagger \xi^{(j,j')}$  in the following way. For  $\ell = 1, \dots, m$ , let  $\nu_\ell \in \mathbb{R}^n$  be the  $\ell$ 'th row of  $\Phi_G$ . Note that  $\nu_\ell$  can be written as

$$\nu_\ell = 4((r_1^{(\ell)})^2, (r_2^{(\ell)})^2, \dots, (r_n^{(\ell)})^2), \quad (\text{C.18})$$

which implies that  $\|\nu_\ell\|_2 \leq 4\sqrt{n}$ . We can then write  $\Phi_G^\dagger \xi^{(j,j')} = \sum_{\ell=1}^m \nu_\ell (\xi^{(j,j')})_\ell$ , which can be bounded by:

$$\begin{aligned} \|\Phi_G^\dagger \xi^{(j,j')}\|_2 &\leq \sum_{\ell=1}^m \|\nu_\ell\|_2 |(\xi^{(j,j')})_\ell| \\ &\leq 4\sqrt{n} \|\xi^{(j,j')}\|_1 \\ &\leq 4\sqrt{nm} \|\xi^{(j,j')}\|_2 \\ &\leq 16m\sqrt{n}. \end{aligned} \quad (\text{C.19})$$

We can then upper-bound  $\|\text{uvec}(\Phi_D^\dagger \Phi(\text{diag}(\mathbf{u})))\|_\infty$  using a similar approach as before, i.e., using a union bound over all  $1 \leq j < j' \leq n$ , and a Hoeffding-type large deviation bound. This yields:

$$\begin{aligned} \|\text{uvec}(\Phi_D^\dagger \Phi(\text{diag}(\mathbf{u})))\|_\infty \\ \leq 16m\sqrt{n}(1 + \sqrt{2})\epsilon_1''' \sqrt{\frac{\ln(n)}{c'}}, \end{aligned} \quad (\text{C.20})$$

with failure probability at most  $\frac{1}{2}e \cdot n^{-(1+2\sqrt{2})}$ . Finally, we combine (C.10), (C.16) and (C.20) to prove (V.15), as desired.

Next, we will prove (V.16), using the same approach we used to prove (V.15). We will bound  $\|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{z})\|_\infty$ , by splitting it into two terms, using (V.17), as follows:

$$\begin{aligned} \|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{z})\|_\infty \\ \leq \|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{v})\|_\infty \\ + \|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\Phi(\text{diag}(\mathbf{u})))\|_\infty. \end{aligned} \quad (\text{C.21})$$

We can bound the first term in (C.21) as follows:

$$\begin{aligned} \|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{v})\|_\infty \\ = \max_{(j,j') \in T^c} |\mathbf{e}_j^T (\Phi_{D,T^c}^\dagger (I - P)\mathbf{v}) \mathbf{e}_{j'}| \\ \leq \max_{j < j'} |\mathbf{v}^T (I - P)\Phi_D(\mathbf{e}_j \mathbf{e}_{j'}^T)| \\ = \max_{j < j'} |\mathbf{v}^T (I - P)\xi^{(j,j')}|, \end{aligned} \quad (\text{C.22})$$

where we defined  $\xi^{(j,j')}$  previously in equation (C.12). Note that

$$\|(I - P)\xi^{(j,j')}\|_2 \leq \|\xi^{(j,j')}\|_2 \leq 4\sqrt{m}, \quad (\text{C.23})$$

since  $P$  is a projector. We can then repeat the same argument that led to equation (C.16). This yields:

$$\|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\mathbf{v})\|_\infty \leq 4\sqrt{m}(1 + \sqrt{2})\epsilon_2''' \sqrt{\frac{\ln(n)}{c'}}, \quad (\text{C.24})$$

with failure probability at most  $\frac{1}{2}e \cdot n^{-(1+2\sqrt{2})}$ .

We now bound the second term in (C.21), using a similar approach:

$$\begin{aligned} \|\text{uvec}(\Phi_{D,T^c}^\dagger (I - P)\Phi(\text{diag}(\mathbf{u})))\|_\infty \\ = \max_{(j,j') \in T^c} |\mathbf{e}_j^T (\Phi_{D,T^c}^\dagger (I - P)\Phi(\text{diag}(\mathbf{u}))) \mathbf{e}_{j'}| \\ \leq \max_{j < j'} |\mathbf{u}^T \Phi_G^\dagger (I - P)\Phi_D(\mathbf{e}_j \mathbf{e}_{j'}^T)| \\ = \max_{j < j'} |\mathbf{u}^T \Phi_G^\dagger (I - P)\xi^{(j,j')}|. \end{aligned} \quad (\text{C.25})$$



We can bound  $\Phi_G^\dagger(I - P)\xi^{(j,j')}$  in the same way as before:

$$\begin{aligned} \|\Phi_G^\dagger(I - P)\xi^{(j,j')}\|_2 &\leq \sum_{\ell=1}^m \|\nu_\ell\|_2 |((I - P)\xi^{(j,j')})_\ell| \\ &\leq 4\sqrt{n} \|(I - P)\xi^{(j,j')}\|_1 \\ &\leq 4\sqrt{nm} \|(I - P)\xi^{(j,j')}\|_2 \\ &\leq 16m\sqrt{n}. \end{aligned} \quad (\text{C.26})$$

We can then repeat the same argument that led to equation (C.20). This yields:

$$\begin{aligned} \|\text{uvec}(\Phi_{D,T^c}^\dagger(I - P)\Phi(\text{diag}(\mathbf{u})))\|_\infty \\ \leq 16m\sqrt{n}(1 + \sqrt{2})\epsilon_1''' \sqrt{\ln(n)/c'}, \end{aligned} \quad (\text{C.27})$$

with failure probability at most  $\frac{1}{2}e \cdot n^{-(1+2\sqrt{2})}$ . Finally, we combine (C.21), (C.24) and (C.27) to prove (V.16), as desired.

#### Appendix D: Convergence of the Random Walk

Here we prove equation (VII.6), which shows the convergence of the random walk that is used to choose the evolution time  $t$ .

We will bound the upper tail probability,  $\Pr[s_{N_{\text{steps}}} \geq s^* + \ell \mid s_0 = 0]$ . (The lower tail,  $\Pr[s_{N_{\text{steps}}} \leq s^* - \ell \mid s_0 = 0]$ , can be handled in a similar way.)

The basic idea is to argue that, if  $s_{N_{\text{steps}}} \geq s^* + \ell$ , then the random walk must spend a significant amount of time in the region above  $s^*$ , and this is unlikely to happen, because the walk tends to drift down towards  $s^*$ . More precisely, let us define the events

$$E_{-1} = \left( \bigwedge_{k=0}^{N_{\text{steps}}-1} (s_k \geq s^* + 1) \right) \wedge (s_{N_{\text{steps}}} \geq s^* + \ell), \quad (\text{D.1})$$

and for  $0 \leq j \leq N_{\text{steps}} - 1$ ,

$$\begin{aligned} E_j = (s_j = s^*) \wedge \left( \bigwedge_{k=j+1}^{N_{\text{steps}}-1} (s_k \geq s^* + 1) \right) \\ \wedge (s_{N_{\text{steps}}} \geq s^* + \ell). \end{aligned} \quad (\text{D.2})$$

That is, the event  $E_{-1}$  occurs if the walk is strictly above  $s^*$  at all times  $\geq 0$ , and the walk ends at or above  $s^* + \ell$  at time  $N_{\text{steps}}$ . The event  $E_j$  occurs if the walk touches  $s^*$  at time  $j$ , the walk remains

strictly above  $s^*$  after time  $j$ , and the walk ends at or above  $s^* + \ell$  at time  $N_{\text{steps}}$ . (Note that the event  $E_j$  can only occur if  $j \leq N_{\text{steps}} - \ell$ , since it takes at least  $\ell$  steps for the walk to move from  $s^*$  to  $s^* + \ell$ .) We can bound the upper tail probability in terms of the events  $E_{-1}$  and  $E_j$ , as follows:

$$\begin{aligned} \Pr[s_{N_{\text{steps}}} \geq s^* + \ell \mid s_0 = 0] \\ \leq \Pr[E_{-1} \mid s_0 = 0] + \sum_{j=0}^{N_{\text{steps}}-\ell} \Pr[E_j \mid s_0 = 0]. \end{aligned} \quad (\text{D.3})$$

(Note that some of these events may occur with probability 0. For instance, the event  $E_0$  (conditioned on  $s_0 = 0$ ) can only occur if  $s^* = 0$ . Similarly, the event  $E_{-1}$  (conditioned on  $s_0 = 0$ ) can only occur if  $s^* \leq -1$ .)

We now use martingale techniques to bound the probabilities of the events  $E_{-1}$  and  $E_j$ . For  $j \geq 1$ , let us define random variables

$$z_j = s_j - \mathbb{E}[s_j \mid s_{j-1}, \dots, s_0]. \quad (\text{D.4})$$

Note that the  $z_j$  form a martingale difference sequence with respect to the  $s_j$ , that is,  $\mathbb{E}[z_j \mid s_{j-1}, \dots, s_0] = 0$ , and the  $z_j$  are bounded,  $|z_j| \leq 2$ . In addition, for  $j' \geq 1$ , let us define random variables

$$\gamma_{j'} = \sum_{j=1}^{j'} z_j. \quad (\text{D.5})$$

(We also define  $\gamma_0 = 0$ .) Note that the  $\gamma_j$  form a martingale with respect to the  $s_j$ , that is,  $\mathbb{E}[\gamma_j \mid s_{j-1}, \dots, s_0] = \gamma_{j-1}$ , and the  $\gamma_j$  have bounded differences,  $|\gamma_j - \gamma_{j-1}| \leq 2$ .

Using these definitions, we can write the following expression for  $s_{j''} - s_{j'}$  (for any  $j'' \geq j'$ ):

$$\begin{aligned} s_{j''} - s_{j'} &= \sum_{k=j'+1}^{j''} z_k + \mathbb{E}[s_k \mid s_{k-1}, \dots, s_0] - s_{k-1} \\ &= \gamma_{j''} - \gamma_{j'} + \sum_{k=j'+1}^{j''} \mathbb{E}[s_k \mid s_{k-1}, \dots, s_0] - s_{k-1}. \end{aligned} \quad (\text{D.6})$$

Furthermore, via Azuma's inequality, we have the following large deviation bound: for any  $\lambda \geq 0$ ,

$$\Pr[\gamma_{j''} - \gamma_{j'} \geq \lambda] \leq \exp\left(-\frac{\lambda^2}{8(j''-j')}\right). \quad (\text{D.7})$$

We will now bound the probability  $\Pr[E_{-1} | s_0 = 0]$ . As noted earlier, this event implies that  $s^* \leq -1$ . We know that  $s^* \geq -h$ , hence  $s_0 \leq s^* + h$ . Using equations (VII.3), (D.1) and (D.6), this implies:

$$\begin{aligned} \ell - h &\leq s_{N_{\text{steps}}} - s_0 \\ &\leq \gamma_{N_{\text{steps}}} - \gamma_0 - \mu N_{\text{steps}}. \end{aligned} \quad (\text{D.8})$$

Hence, using (D.7), and the fact that  $N_{\text{steps}} = \frac{h}{\mu} + \eta$ , we have:

$$\begin{aligned} &\Pr[E_{-1} | s_0 = 0] \\ &\leq \Pr[\gamma_{N_{\text{steps}}} - \gamma_0 \geq \ell - h + \mu N_{\text{steps}} | s_0 = 0] \\ &\leq \exp\left(-\frac{(\ell - h + \mu N_{\text{steps}})^2}{8N_{\text{steps}}}\right) \\ &= \exp\left(-\frac{(\ell + \mu\eta)^2}{8\left(\frac{h}{\mu} + \eta\right)}\right) \\ &\leq \exp\left(-\frac{\mu\eta}{16 \max\{\frac{h}{\mu}, \eta\}}(\ell + \mu\eta)\right) \\ &= \exp\left(-\frac{\mu}{16} \min\{\frac{\mu\eta}{h}, 1\}(\ell + \mu\eta)\right). \end{aligned} \quad (\text{D.9})$$

In a similar way, we can bound the probability

$\Pr[E_j | s_0 = 0]$ . Suppose that the event  $E_j$  happens. Using equations (VII.3), (D.2) and (D.6), this implies:

$$\begin{aligned} \ell &\leq s_{N_{\text{steps}}} - s_j \\ &\leq \gamma_{N_{\text{steps}}} - \gamma_j - \mu(N_{\text{steps}} - j - 1). \end{aligned} \quad (\text{D.10})$$

Hence, using (D.7), and the fact that  $\ell \geq 1 \geq \mu$ , we have:

$$\begin{aligned} &\Pr[E_j | s_0 = 0] \\ &\leq \Pr[\gamma_{N_{\text{steps}}} - \gamma_j \geq \ell + \mu(N_{\text{steps}} - j - 1) | s_0 = 0] \\ &\leq \exp\left(-\frac{(\ell + \mu(N_{\text{steps}} - j - 1))^2}{8(N_{\text{steps}} - j)}\right) \\ &\leq \exp\left(-\frac{\mu}{8}(\ell + \mu(N_{\text{steps}} - j - 1))\right). \end{aligned} \quad (\text{D.11})$$

We can now bound the sum of these probabilities as follows:

$$\sum_{j=0}^{N_{\text{steps}}-\ell} \Pr[E_j | s_0 = 0] \leq \sum_{j=0}^{N_{\text{steps}}-\ell} \exp\left(-\frac{\mu}{8}(\ell + \mu(N_{\text{steps}} - j - 1))\right) \quad (\text{D.12})$$

$$= \exp\left(-\frac{\mu}{8}\ell - \frac{\mu^2}{8}(N_{\text{steps}} - 1)\right) \sum_{j=0}^{N_{\text{steps}}-\ell} \exp\left(\frac{\mu^2}{8}j\right) \quad (\text{D.13})$$

$$= \exp\left(-\frac{\mu}{8}\ell - \frac{\mu^2}{8}(N_{\text{steps}} - 1)\right) \frac{\exp\left(\frac{\mu^2}{8}(N_{\text{steps}} - \ell + 1)\right) - 1}{\exp\left(\frac{\mu^2}{8}\right) - 1} \quad (\text{D.14})$$

$$\leq \exp\left(-\frac{\mu}{8}\ell - \frac{\mu^2}{8}(N_{\text{steps}} - 1)\right) \frac{8}{\mu^2} \exp\left(\frac{\mu^2}{8}(N_{\text{steps}} - \ell + 1)\right) \quad (\text{D.15})$$

$$\leq \frac{8}{\mu^2} \exp\left(-\frac{\mu(\mu+1)}{8}\ell + \frac{\mu^2}{4}\right). \quad (\text{D.16})$$

Finally, combining equations (D.3), (D.9) and (D.12), we get the desired bound on the upper tail  $\Pr[s_{N_{\text{steps}}} \geq s^* + \ell | s_0 = 0]$ . The bound on the lower tail  $\Pr[s_{N_{\text{steps}}} \leq s^* - \ell | s_0 = 0]$  is proved in a similar way. This completes the proof of (VII.6).

## Appendix E: Bounding SPAM errors

We first give a derivation of equations (VIII.7) and (VIII.8). Note that the evolution of  $E_0$  and  $\rho_0$  is known. Additionally, we have  $\text{Tr}[E_0 \mathcal{E}_t(\delta\rho)] = \text{Tr}[\mathcal{E}_t^\dagger(E_0)\delta\rho]$ , where  $\mathcal{E}_t^\dagger = \exp(\mathcal{L}^\dagger t)$  is the adjoint dephasing map, which coincides with  $\mathcal{E}_t$  in our case as

the Lindblad operators are all Hermitian. Therefore,

$$\text{Tr}[E_0\mathcal{E}_t(\delta\rho)] = \text{Tr}[\mathcal{E}_t^\dagger(|\mathbf{a}\rangle\langle\mathbf{a}|)\delta\rho] + \text{Tr}[\mathcal{E}_t^\dagger(|\mathbf{b}\rangle\langle\mathbf{b}|)\delta\rho] + \text{Tr}[\mathcal{E}_t^\dagger(|\mathbf{a}\rangle\langle\mathbf{b}|)\delta\rho] + \text{Tr}[\mathcal{E}_t^\dagger(|\mathbf{b}\rangle\langle\mathbf{a}|)\delta\rho] \quad (\text{E.1})$$

$$= \text{Tr}[|\mathbf{a}\rangle\langle\mathbf{a}|\delta\rho] + \text{Tr}[|\mathbf{b}\rangle\langle\mathbf{b}|\delta\rho] + e^{-\Gamma_{\mathbf{r}}t}\text{Tr}[|\mathbf{a}\rangle\langle\mathbf{b}|\delta\rho] + e^{-\Gamma_{\mathbf{r}}t}\text{Tr}[|\mathbf{b}\rangle\langle\mathbf{a}|\delta\rho] \quad (\text{E.2})$$

$$= \eta_s + \zeta_s e^{-\Gamma_{\mathbf{ab}}t} \quad (\text{E.3})$$

$$\text{Tr}[\delta E\mathcal{E}_t(\rho_0)] = \text{Tr}[\delta E\mathcal{E}_t(|\mathbf{a}\rangle\langle\mathbf{a}|)] + \text{Tr}[\delta E\mathcal{E}_t(|\mathbf{b}\rangle\langle\mathbf{b}|)] + \text{Tr}[\delta E\mathcal{E}_t(|\mathbf{a}\rangle\langle\mathbf{b}|)] + \text{Tr}[\delta E\mathcal{E}_t(|\mathbf{b}\rangle\langle\mathbf{a}|)] \quad (\text{E.4})$$

$$= \text{Tr}[\delta E|\mathbf{a}\rangle\langle\mathbf{a}|] + \text{Tr}[\delta E|\mathbf{b}\rangle\langle\mathbf{b}|] + e^{-\Gamma_{\mathbf{r}}t}\text{Tr}[\delta E|\mathbf{a}\rangle\langle\mathbf{b}|] + e^{-\Gamma_{\mathbf{r}}t}\text{Tr}[\delta E|\mathbf{b}\rangle\langle\mathbf{a}|] \quad (\text{E.5})$$

$$= \eta_m + \zeta_m e^{-\Gamma_{\mathbf{ab}}t} \quad (\text{E.6})$$

We see that both terms decay with the rate  $\Gamma_{\mathbf{ab}}$ . Therefore, as noted in Section VIII, the outcome of the protocol in the presence of error is given by  $\tilde{P}_{\mathbf{ab}}(t) = \frac{1}{2}(1 + \eta_s + \eta_m + (1 + \zeta_s + \zeta_m)e^{-\Gamma_{\mathbf{ab}}t}) + R(t)$ , where deviations form a single exponential decay  $e^{-\Gamma_{\mathbf{ab}}t}$  are captured in  $R(t) = \text{Tr}[\delta E\mathcal{E}_t(\delta\rho)]$ . To bound the growth of this term we have

$$|\dot{R}(t)| = \left| \frac{\partial}{\partial t} \text{Tr}[\delta E\mathcal{E}_t(\delta\rho)] \right| \quad (\text{E.7})$$

$$= |\text{Tr}[\delta E\mathcal{L}(\delta\rho)]| \quad (\text{E.8})$$

$$\leq \|\delta E\| \|\mathcal{L}(\delta\rho)\|_{\text{tr}} \quad (\text{E.9})$$

$$\leq \|\delta E\| \left\| \sum_{ij} L_i \delta\rho L_j^\dagger - \frac{1}{2} \{L_j^\dagger L_i, \delta\rho\} \right\|_{\text{tr}} \quad (\text{E.10})$$

$$\leq \|\delta E\| \left\| \sum_{ij} \|L_i \delta\rho L_j^\dagger - \frac{1}{2} \{L_j^\dagger L_i, \delta\rho\} \right\|_{\text{tr}} \quad (\text{E.11})$$

$$\leq \|\delta E\| \|\delta\rho\|_{\text{tr}} \sum_{ij} 2\|L_i\| \|L_j^\dagger\| \quad (\text{E.12})$$

$$\leq 2\epsilon_m \epsilon_s (n + s) \quad (\text{E.13})$$

where we used the Hölder's inequality [54] in going from (E.8) to (E.9). Additionally, in deriving (E.12) from (E.11) we used the fact that  $\|AB\|_{\text{tr}} \leq \|A\| \|B\|_{\text{tr}}$ , which we will prove in the following. Finally, the last line is obtained by noting that  $L_i = Z_i$  in our problem and  $\|Z_i\| = 1$ .

We can obtain  $\|AB\|_{\text{tr}} \leq \|A\| \|B\|_{\text{tr}}$ , by first showing that for any matrix  $A$  and  $B$ ,  $\sigma(AB)_i \leq \sigma_{\max}(A)\sigma_i(B)$ , where  $\sigma_i$  are the ordered singular val-

ues [55]. We then use that to obtain

$$\|AB\|_{\text{tr}} = \sum_i \sigma_i(AB) \quad (\text{E.14})$$

$$\leq \sigma_{\max}(B) \sum_i \sigma_i(A) \quad (\text{E.15})$$

$$= \|B\| \|A\|_{\text{tr}}. \quad (\text{E.16})$$

To show that  $\sigma(AB)_i \leq \sigma_{\max}(A)\sigma_i(B)$ , first note that for Hermitian matrices  $A$  and  $B$  we have:

$$A \geq B \implies \lambda_i(A) \geq \lambda_i(B) \quad (\text{E.17})$$

where  $\lambda_i$ 's are ordered eigenvalues. We then show that  $\lambda_i(AB) \leq \lambda_{\max}(A)\lambda_i(B)$  if  $A \geq 0$  and  $B \geq 0$ . Note that  $\lambda_i(AB) = \lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}})$ , which can be seen by

$$AB|i\rangle = \lambda_i|i\rangle \implies B^{\frac{1}{2}}AB^{\frac{1}{2}}(B^{\frac{1}{2}}|i\rangle) = \lambda_i(B^{\frac{1}{2}}|i\rangle) \quad (\text{E.18})$$

Next, note that  $\lambda_{\max}(A)I - A \geq 0$ . Therefore,  $B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}} \geq 0$ . This is because  $\langle x|B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}}|x\rangle = \langle y|\lambda_{\max}(A)I - A|y\rangle \geq 0$  for all  $|y\rangle = B^{\frac{1}{2}}|x\rangle$ . We then obtain

$$\begin{aligned} B^{\frac{1}{2}}AB^{\frac{1}{2}} &\leq B^{\frac{1}{2}}AB^{\frac{1}{2}} + B^{\frac{1}{2}}(\lambda_{\max}(A)I - A)B^{\frac{1}{2}} \\ &= \lambda_{\max}(A)B \end{aligned} \quad (\text{E.19})$$

Therefore,  $B^{\frac{1}{2}}AB^{\frac{1}{2}} \leq \lambda_{\max}(A)B$  together with (E.17) implies that

$$\lambda_i(AB) = \lambda_i(B^{\frac{1}{2}}AB^{\frac{1}{2}}) \leq \lambda_{\max}(A)\lambda_i(B). \quad (\text{E.20})$$

Now for arbitrary square matrices  $A$  and  $B$  we have

$$\sigma_i(AB) = \sqrt{\lambda_i(B^\dagger A^\dagger AB)} = \sqrt{\lambda_i(A^\dagger A B B^\dagger)}. \quad (\text{E.21})$$

Therefore, we can use (E.20) for positive-semidefinite matrices  $A^\dagger A$  and  $BB^\dagger$  to obtain

$$\lambda_i(A^\dagger A B B^\dagger) \leq \lambda_{\max}(A^\dagger A) \lambda_i(B B^\dagger). \quad (\text{E.22})$$

Since both sides of the inequality are positive, we can take their square root to obtain

$$\sigma_i(AB) \leq \sigma_{\max}(A) \sigma_i(B). \quad (\text{E.23})$$

## Appendix F: Isotropy of the measurements $\Omega_{\text{ab}}$

In this section we show that the random vectors  $\mathbf{q}$  in equation (IX.11) are centered and isotropic (up to a normalization factor of  $\sqrt{2}$ ), as claimed in equation (IX.12). Noting that  $\alpha_i$  and  $\beta_i$  are chosen uniformly and independently at random from  $\{-1, 1\}$  we find that

$$\mathbb{E}[\alpha_i \alpha_j] = \mathbb{E}[\alpha_i \beta_j] = \mathbb{E}[\beta_i \beta_j] = 0, \quad (\text{F.1})$$

which shows that  $\mathbb{E}[\mathbf{q}] = 0$ .

To show that  $\mathbb{E}(\mathbf{q}\mathbf{q}^T) = I$ , we have to consider three types of correlations: (i) correlations between the elements of  $\text{uvec}(\boldsymbol{\alpha}\boldsymbol{\beta}^T - \boldsymbol{\beta}\boldsymbol{\alpha}^T)$ , (ii) correlations between the elements of  $\text{uvec}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \boldsymbol{\beta}\boldsymbol{\beta}^T)$ , and (iii) cross-correlations between these two types of elements.

We start with correlations of type (i). We consider

$$\begin{aligned} & \mathbb{E}[(\alpha_i \beta_j - \beta_i \alpha_j)(\alpha_k \beta_l - \beta_k \alpha_l)] \\ &= \mathbb{E}[\alpha_i \beta_j \alpha_k \beta_l - \alpha_i \beta_j \beta_k \alpha_l - \beta_i \alpha_j \alpha_k \beta_l + \beta_i \alpha_j \beta_k \alpha_l]. \end{aligned} \quad (\text{F.2})$$

Similar to the analysis in Section VD, we consider three different cases for how the sets of indices can intersect. Note that in the following we assume  $i < j$  and  $k < l$ , as we are only considering the off-diagonal upper-triangular parts of  $V$ . If  $\{i, j\} \cap \{k, l\} = \emptyset$  we have

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \beta_l - \alpha_i \beta_j \beta_k \alpha_l - \beta_i \alpha_j \alpha_k \beta_l + \beta_i \alpha_j \beta_k \alpha_l] = 0, \quad (\text{F.3})$$

because of the symmetry of the terms. If  $|\{i, j\} \cap \{k, l\}| = 1$ , without the loss of generality we assume that  $i = k$  and find

$$\begin{cases} \mathbb{E}[\alpha_i \beta_j \alpha_k \beta_l] &= \mathbb{E}[\alpha_i^2 \beta_j \beta_l] = \mathbb{E}[\beta_j \beta_l] = 0 \\ \mathbb{E}[\alpha_i \beta_j \beta_k \alpha_l] &= \mathbb{E}[\alpha_i \beta_i \beta_j \beta_l] = 0 \\ \mathbb{E}[\beta_i \alpha_j \alpha_k \beta_l] &= \mathbb{E}[\beta_i \alpha_i \alpha_j \beta_l] = 0 \\ \mathbb{E}[\beta_i \alpha_j \beta_k \alpha_l] &= \mathbb{E}[\beta_i^2 \alpha_j \alpha_l] = \mathbb{E}[\alpha_j \alpha_l] = 0 \end{cases} \quad (\text{F.4})$$

Therefore, we find

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \beta_l - \alpha_i \beta_j \beta_k \alpha_l - \beta_i \alpha_j \alpha_k \beta_l + \beta_i \alpha_j \beta_k \alpha_l] = 0. \quad (\text{F.5})$$

Note that if we had assumed  $j = l$  we would have obtained the same results. Finally, when  $|\{i, j\} \cap \{k, l\}| = 2$ , we have that  $i = k$  and  $j = l$  and find

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \beta_l - \alpha_i \beta_j \beta_k \alpha_l - \beta_i \alpha_j \alpha_k \beta_l + \beta_i \alpha_j \beta_k \alpha_l] \quad (\text{F.6})$$

$$= \mathbb{E}[\alpha_i^2 \beta_j^2 - \alpha_i \beta_i \alpha_j \beta_j - \alpha_i \beta_i \alpha_j \beta_j + \beta_i^2 \alpha_j^2] \quad (\text{F.7})$$

$$= 2 \quad (\text{F.8})$$

Next we consider correlations of type (ii), that is,

$$\begin{aligned} & \mathbb{E}[(\alpha_i \alpha_j - \beta_i \beta_j)(\alpha_k \alpha_l - \beta_k \beta_l)] \\ &= \mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l], \end{aligned} \quad (\text{F.9})$$

where  $i < j$  and  $k < l$ . Similarly, we consider different cases: If  $\{i, j\} \cap \{k, l\} = \emptyset$  we have

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] = 0, \quad (\text{F.10})$$

because of the symmetries. If  $|\{i, j\} \cap \{k, l\}| = 1$ , without the loss of generality we assume that  $i = k$  and find

$$\begin{cases} \mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l] &= \mathbb{E}[\alpha_i^2 \alpha_j \alpha_l] = \mathbb{E}[\alpha_j \alpha_l] = 0 \\ \mathbb{E}[\alpha_i \alpha_j \beta_k \beta_l] &= \mathbb{E}[\alpha_i \beta_i \alpha_j \beta_l] = 0 \\ \mathbb{E}[\beta_i \beta_j \alpha_k \alpha_l] &= \mathbb{E}[\beta_i \alpha_i \beta_j \alpha_l] = 0 \\ \mathbb{E}[\beta_i \beta_j \beta_k \beta_l] &= \mathbb{E}[\beta_i^2 \beta_j \beta_l] = \mathbb{E}[\beta_j \beta_l] = 0 \end{cases} \quad (\text{F.11})$$

Therefore, we find

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] = 0. \quad (\text{F.12})$$

Finally, when  $|\{i, j\} \cap \{k, l\}| = 2$ , we have  $i = k$  and  $j = l$  and find

$$\mathbb{E}[\alpha_i \alpha_j \alpha_k \alpha_l - \alpha_i \alpha_j \beta_k \beta_l - \beta_i \beta_j \alpha_k \alpha_l + \beta_i \beta_j \beta_k \beta_l] \quad (\text{F.13})$$

$$= \mathbb{E}[\alpha_i^2 \alpha_j^2 - \alpha_i \beta_i \alpha_j \beta_j - \alpha_i \beta_i \alpha_j \beta_j + \beta_i^2 \beta_j^2] \quad (\text{F.14})$$

$$= 2. \quad (\text{F.15})$$

Finally, we consider cross-correlations of type (iii), which are given by

$$\begin{aligned} & \mathbb{E}[(\alpha_i \beta_j - \beta_i \alpha_j)(\alpha_k \alpha_l - \beta_k \beta_l)] \\ &= \mathbb{E}[\alpha_i \beta_j \alpha_k \alpha_l - \alpha_i \beta_j \beta_k \beta_l - \beta_i \alpha_j \alpha_k \alpha_l + \beta_i \alpha_j \beta_k \beta_l], \end{aligned} \quad (\text{F.16})$$

with  $i < j$  and  $k < l$ . If  $\{i, j\} \cap \{k, l\} = \emptyset$  we have

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \alpha_l - \alpha_i \beta_j \beta_k \beta_l - \beta_i \alpha_j \alpha_k \alpha_l + \beta_i \alpha_j \beta_k \beta_l] = 0, \quad (\text{F.17})$$

because of the symmetries. If  $|\{i, j\} \cap \{k, l\}| = 1$ , without the loss of generality we assume that  $i = k$  and find

$$\begin{cases} \mathbb{E}[\alpha_i \beta_j \alpha_k \alpha_l] &= \mathbb{E}[\alpha_i^2 \beta_j \alpha_l] = \mathbb{E}[\beta_j \alpha_l] = 0 \\ \mathbb{E}[\alpha_i \beta_j \beta_k \beta_l] &= \mathbb{E}[\alpha_i \beta_j \beta_i \beta_l] = 0 \\ \mathbb{E}[\beta_i \alpha_j \alpha_k \alpha_l] &= \mathbb{E}[\beta_i \alpha_j \alpha_i \alpha_l] = 0 \\ \mathbb{E}[\beta_i \alpha_j \beta_k \beta_l] &= \mathbb{E}[\beta_i^2 \alpha_j \beta_l] = \mathbb{E}[\alpha_j \beta_l] = 0 \end{cases}. \quad (\text{F.18})$$

Therefore, we find

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \alpha_l - \alpha_i \beta_j \beta_k \beta_l - \beta_i \alpha_j \alpha_k \alpha_l + \beta_i \alpha_j \beta_k \beta_l] = 0. \quad (\text{F.19})$$

Finally, when  $|\{i, j\} \cap \{k, l\}| = 2$ , we assume  $i = k$  and  $j = l$  and find

$$\mathbb{E}[\alpha_i \beta_j \alpha_k \alpha_l - \alpha_i \beta_j \beta_k \beta_l - \beta_i \alpha_j \alpha_k \alpha_l + \beta_i \alpha_j \beta_k \beta_l] \quad (\text{F.20})$$

$$= \mathbb{E}[\alpha_i^2 \beta_j \alpha_j - \alpha_i \beta_i \beta_j^2 - \beta_i \alpha_i \alpha_j^2 + \beta_i^2 \alpha_j \beta_j] \quad (\text{F.21})$$

$$= 0 \quad (\text{F.22})$$

Therefore, given all cases considered here we arrive at equation (IX.12).

- 
- [1] J. Preskill, *Quantum* **2**, 79 (2018).
- [2] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, arXiv preprint arXiv:1910.06343 (2019).
- [3] E. J. Candes and M. B. Wakin, *IEEE Signal Processing Magazine* **25**, 21 (2008).
- [4] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Physical Review Letters* **105**, 150401 (2010).
- [5] A. Shabani, R. Kosut, M. Mohseni, H. Rabitz, M. Broome, M. Almeida, A. Fedrizzi, and A. White, *Physical Review Letters* **106**, 100401 (2011).
- [6] A. Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, *Physical Review A* **84**, 012107 (2011).
- [7] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, *New Journal of Physics* **14**, 095022 (2012).
- [8] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, *Phys. Rev. Lett.* **121**, 170502 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.170502>.
- [9] Y. Liu, in *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger (2011), pp. 1638–1646, URL <http://papers.nips.cc/paper/4222-universal-low-rank-matrix-recovery-from-pauli-observables>.
- [10] R. Harper, W. Yu, and S. T. Flammia, *PRX Quantum* **2**, 010322 (2021), URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.010322>.
- [11] H.-P. Breuer, F. Petruccione, et al., *The theory of open quantum systems* (Oxford University Press on Demand, 2002).
- [12] G. M. Palma, K.-A. Suominen, and A. Ekert, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **452**, 567 (1996).
- [13] L.-M. Duan and G.-C. Guo, *Physical Review A* **57**, 737 (1998).
- [14] J. H. Reina, L. Quiroga, and N. F. Johnson, *Physical Review A* **65**, 032326 (2002).
- [15] R. Klesse and S. Frank, *Phys. Rev. Lett.* **95**, 230503 (2005), URL <https://link.aps.org/doi/10.1103/PhysRevLett.95.230503>.
- [16] D. Layden and P. Cappellaro, *npj Quantum Information* **4**, 30 (2018).
- [17] D. Layden, S. Zhou, P. Cappellaro, and L. Jiang, *Phys. Rev. Lett.* **122**, 040502 (2019), URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.040502>.
- [18] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Quantum* **4**, 321 (2020).
- [19] R. Harper, S. T. Flammia, and J. J. Wallman, *Nature Physics* pp. 1–5 (10 August 2020).
- [20] L. Govia, G. Ribeill, D. Ristè, M. Ware, and H. Krovi, *Nature Communications* **11**, 1 (2020).
- [21] J. Jeske and J. H. Cole, *Physical Review A* **87**, 052138 (2013).
- [22] N. M. Linke, D. Maslov, M. Roetteler, S. Deb-nath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, *Proceedings of the National Academy of Sciences* **114**, 3305 (2017).
- [23] R. Beals, S. Brierley, O. Gray, A. W. Harrow, S. Kutin, N. Linden, D. Shepherd, and M. Stather, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **469**, 20120686 (2013).
- [24] N. F. Ramsey, *Phys. Rev.* **78**, 695 (1950), URL <https://link.aps.org/doi/10.1103/PhysRev.78.695>.
- [25] C. L. Degen, F. Reinhard, and P. Cappellaro, *Rev.*

- Mod. Phys. **89**, 035002 (2017), URL <https://link.aps.org/doi/10.1103/RevModPhys.89.035002>.
- [26] S. S. Chen, D. L. Donoho, and M. A. Saunders, SIAM Review **43**, 129 (2001).
- [27] R. Vershynin, in *Compressed Sensing: Theory and Applications*, edited by Y. C. Eldar and G. Kutyniok (Cambridge University Press, 2012), p. 210–268.
- [28] E. J. Candes and T. Tao, IEEE Transactions on Information Theory **52**, 5406 (2006).
- [29] M. Rudelson and R. Vershynin, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **61**, 1025 (2008).
- [30] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Applied and Numerical Harmonic Analysis (Springer New York, 2013), ISBN 9780817649487, URL <https://books.google.com/books?id=zb28BAAAQBAJ>.
- [31] E. J. Candes, J. K. Romberg, and T. Tao, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **59**, 1207 (2006).
- [32] E. J. Candes and Y. Plan, IEEE Transactions on Information Theory **57**, 7235 (2011).
- [33] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, in *Compressed Sensing: Theory and Applications*, edited by Y. C. Eldar and G. Kutyniok (Cambridge University Press, 2012), pp. 1–64.
- [34] E. J. Candes, T. Strohmer, and V. Voroninski, Communications on Pure and Applied Mathematics **66**, 1241 (2013).
- [35] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012), vol. 25, URL <https://proceedings.neurips.cc/paper/2012/file/3a066bda8c96b9478bb0512f0a43028c-Paper.pdf>.
- [36] X. Li and V. Voroninski, SIAM Journal on Mathematical Analysis **45**, 3019 (2013).
- [37] K. Jaganathan, S. Oymak, and B. Hassibi, in *2013 IEEE International Symposium on Information Theory (IEEE, 2013)*, pp. 1022–1026.
- [38] R. Kueng, H. Rauhut, and U. Terstiege, Applied and Computational Harmonic Analysis **42**, 88 (2017).
- [39] D. Suess, N. Maraviglia, R. Kueng, A. Mañnos, C. Sparrow, T. Hashimoto, N. Matsuda, D. Gross, and A. Laing, *Rapid characterisation of linear-optical networks via phaselift* (2020), arXiv preprint 2010.00517.
- [40] J. Emerson, R. Alicki, and K. Życzkowski, Journal of Optics B: Quantum and Semiclassical Optics **7**, S347 (2005).
- [41] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008), URL <https://link.aps.org/doi/10.1103/PhysRevA.77.012307>.
- [42] D. Greenbaum, arXiv preprint arXiv:1509.02921 (2015).
- [43] A. Kalev, R. L. Kosut, and I. H. Deutsch, npj Quantum Information **1**, 1 (2015).
- [44] R. Kueng (2021), personal communication.
- [45] T. Sugiyama, P. S. Turner, and M. Muraio, Phys. Rev. Lett. **111**, 160406 (2013), URL <https://link.aps.org/doi/10.1103/PhysRevLett.111.160406>.
- [46] M. Guță, J. Kahn, R. Kueng, and J. A. Tropp, Journal of Physics A: Mathematical and Theoretical **53**, 204001 (2020).
- [47] S. Diamond and S. Boyd, Journal of Machine Learning Research **17**, 1 (2016).
- [48] A. Agrawal, R. Verschuere, S. Diamond, and S. Boyd, Journal of Control and Decision **5**, 42 (2018).
- [49] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics (Springer New York, 2013), ISBN 9780387216065, URL <https://books.google.com/books?id=yPfZBwAAQBAJ>.
- [50] A. P. Thorne, *Spectrophysics* (Springer Netherlands, 1988), URL <https://doi.org/10.1007/978-94-009-1193-2>.
- [51] G. A. Paz-Silva, L. M. Norris, and L. Viola, Physical Review A **95**, 022121 (2017).
- [52] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters **96**, 010401 (2006).
- [53] C. Addis, P. Haikka, S. McEndoo, C. Macchiavello, and S. Maniscalco, Physical Review A **87**, 052109 (2013).
- [54] B. Baumgartner, arXiv preprint arXiv:1106.6189 (2011).
- [55] F. Zhang, *Matrix theory: basic results and techniques* (Springer Science & Business Media, 2011).