Error models in quantum computation: An application of model selection

Lucia Schwarz and S. J. van Enk

Department of Physics, Oregon Center for Optics, University of Oregon, Eugene, Oregon 97403, USA

Threshold theorems for fault-tolerant quantum computing assume that errors are of certain types. But how would one detect whether errors of the "wrong" type occur in one's experiment, especially if one does not even know what type of error to look for? The problem is that for many qubits a full state description is almost impossible to analyze due to the exponentially large state space, and a full process description requires even more resources. As a result, one simply *cannot* detect all types of errors. Here we show through a quantum state estimation example (on up to 25 qubits) how to attack this problem using *model selection*. We use, in particular, the Akaike information criterion. The example indicates that the number of measurements that one has to perform before noticing errors of the wrong type scales polynomially both with the number of qubits and with the error size.

I. INTRODUCTION

In order to develop a quantum computer we need to be able to coherently control and read out a system of many qubits. Verifying how a particular experimental implementation of a quantum computer actually performs will be straightforward once we can run a computation in a fault tolerant manner: we just check whether the answer produced by the computation is correct or not. However, before that time arrives we will need to employ other less conclusive types of tests.

There are two types of generic tests that provide useful even if somewhat nonspecific information about many-qubit systems: multipartite entanglement verification tests [1] and randomized benchmarking [2,3]. In both cases one may detect that something is wrong, but one will not find out what exactly is wrong. Unfortunately, there is no efficient procedure to figure out what exactly is wrong, simply because we cannot efficiently simulate a generic multiqubit quantum process on a classical computer. (For smaller systems quantum tomography can be used, and even there one has to be careful with systematic and other errors [4–6].)

For fault tolerant quantum computation [7] one does need to know not just how large the error probabilities are but also whether they are of the right type. This is because threshold theorems [8–12] need to make explicit use of error models. For example, the calculation of the error threshold may be based on a "local stochastic" error model (for an introduction, see [13]). Errors correlated over a long range may then be disastrous. One mechanism by which such long-range errors might arise is as follows. A laser field's phase and intensity always fluctuate, but, of course, if those fluctuations are always sufficiently small, the errors they cause will be corrected for by quantum error correction. But what if the fluctuations, for just a brief time interval, are large? Then all qubits which happened to have been accessed during that time interval have a much larger probability of error. The problem we consider is how one could notice the presence of such errors.

While there is no systematic and efficient method to solve this problem completely, there is an efficient and well-tested method: model selection [14,15]. This term refers to a well-developed field of (classical) statistics and inference in which the aim is to rank order different (statistical) models, each meant to describe some given process. In the present context,

model selection boils down to this: We design a few-parameter model that describes our predictions of all the processes and errors that occur in our experiment—it may have a few noise parameters with a clear physical meaning, for instance—and compare it with a much larger (but still far from exhaustive) model that includes many (but not all) possible types of errors. As long as the large model contains a number of parameters that scales moderately with the number of qubits, then it still can be analyzed, even for a few dozen qubits. If that large model is ranked higher than our few-parameter model, we conclude that errors occurred that we did not expect.

We are going to discuss an illustrative example of this model selection procedure. We simulate a quantum state estimation experiment on N qubits, which is modeled after an actual experiment performed on 14 ions in an ion trap in which a 14-qubit Greenberger-Horne-Zeilinger (GHZ) state of high fidelity was generated [16]. We will vary N up to 25 and assume the goal is to generate a perfect GHZ state. We take a three-parameter model (with three noise parameters describing three different noise processes) as our standard error model and then take a model with $O(N^3)$ parameters as the much larger error model, which includes many types of errors, although, obviously, not all $O(4^N)$ possible ones. We assume the data are generated by a "true" state of the form

$$\rho_{\text{true}} = (1 - q)\rho_{\text{SE}} + q\rho_{\text{GE}},$$

with the subscript SE referring to the standard error model and GE referring to the more general error model. We investigate then the following issues: First, does the model selection procedure recognize that the standard error model is indeed correct (i.e., ranked higher than the large general error model) when $q = 0$? Second, in the case that $q \neq 0$, how many measurements does one need to take before one notices that there is in fact an error that lies outside the standard error model? The last question splits naturally into two subquestions, namely, how that number scales with the number of qubits and how it scales with q.

The model selection method we use here is based on the Akaike information criterion (AIC) [17]. This method is widely used outside of physics and by now has been applied on various occasions within quantum information theory as well [6,18–21]. Most model selection criteria compare the

goodness of fit of each model while penalizing the number of parameters, thus possibly favoring simpler models. The AIC in particular has a clear meaning since it is derived purely from the principles of information (see Sec. II). It has been found to perform better than the related Bayesian information criterion [15] in quantum state and entanglement estimation [19].

II. PRELIMINARIES

A. Model selection and AIC

Suppose we have taken data and now wish to model the underlying process that generated the data. Our data contain some amount of information about the underlying process, but also statistical fluctuations. How can we determine whether a model is a good description of the underlying process rather than of the statistical fluctuations? In general, models with more parameters will fit the data better but are also more likely to fit to the fluctuations, and models with *too many* parameters are *over* fitting. One method to find a compromise between under- and overfitting was proposed by Akaike [17]. He derived an expression for the estimated Kullback-Leibler divergence between one's model and the true underlying process. The Kullback-Leibler divergence is expressed in terms of two probability distributions for the data, the "true distribution" $\{p_i\}$ and the distribution generated by our model, $\{s_i\}$, as follows:

$$D_{KL}(p||s) = \sum_i p_i \ln \frac{p_i}{s_i}. \tag{1}$$

This is a measure for the distance between the two probability distributions $\{p_i\}$ and $\{s_i\}$. It is also called the *relative entropy* and can be understood as the information that is lost if the model $\{s_i\}$ is used instead of the "real" distribution $\{p_i\}$.

Of course, we do not know the true underlying distribution, but, nonetheless, the Kullback-Leibler divergence can be estimated, as was shown by Akaike, using the observed frequencies. Namely, up to a constant that is the same for all models, he found the divergence to approximately equal

$$\text{AIC} = -2\mathcal{L}_{\max} + 2K. \tag{2}$$

Here K is the number of parameters of the model, and $\mathcal{L}_{\max}$ is its maximum ln likelihood:

$$\mathcal{L}_{\max} = \max_{\{p_k\}} \sum_k f_k \ln p_k, \tag{3}$$

with f_k being the number of times outcome k was observed and p_k being the probability according to the model of obtaining outcome k. Model selection now consists of calculating the AIC for different candidate models, with the lowest score corresponding to the best model. In our context this procedure can distinguish between models that accurately describe the relevant physical (error) processes and models that spend too many parameters on fitting statistical noise. We thereby gain insights into the actual physical processes that cause errors, and we can tell whether or not errors outside our simple model are significant.

B. A three-parameter model for noisy GHZ states

We will simulate an experiment on a noisy GHZ state [22] of N qubits. The ideal GHZ state is a coherent superposition of all qubits in state $|0\rangle$ or all in state $|1\rangle$:

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|00...0\rangle + |11...1\rangle). \tag{4}$$

A high-fidelity version of this state was created in a trapped ion system for 14 ions [16]. The density matrix $\rho_{\text{GHZ}} = |\text{GHZ}\rangle\langle\text{GHZ}|$, written in the standard basis $|00...0\rangle, |00...1\rangle, ..., |11...1\rangle$, has only four nonzero elements which all equal $\frac{1}{2}$. This state is maximally entangled and pure. However, a real quantum system in the laboratory will not be in this perfect state. There might be several effects that act on the qubits during the state preparation and/or storage. As a simple and not unreasonable model we assume just three noise processes, described by three parameters: a small imbalance ε between the populations of $|00..0\rangle$ and $|11..1\rangle$, a systematic phase shift φ of the relative phase between $|00..0\rangle$ and $|11..1\rangle$, and δ, which quantifies the loss of coherence between $|00..0\rangle$ and $|11..1\rangle$ due to random-phase fluctuations. These three processes will create a mixed state with density matrix

$$\rho_{3\text{P}} = \frac{1}{2} \begin{pmatrix} 1+\varepsilon & ... & \delta\sqrt{1-\varepsilon^2}e^{i\varphi} \\ \vdots & & \vdots \\ \delta\sqrt{1-\varepsilon^2}e^{-i\varphi} & ... & 1-\varepsilon \end{pmatrix}. \tag{5}$$

C. A large model: permutationally invariant states

For the comparison with our three-parameter model, we try to find a model that will describe many possible errors and deviations from this simple model. However, we can't model *all* possible errors that may occur. Our goal is therefore to design a model with a fairly large (but still polynomial) number of fitting parameters. If an arbitrary error is affecting our experiment, it will most likely be partially contained in this large model and we will detect it. Of course, this leaves out certain errors that are exactly orthogonal to the large model. To increase the chance of detecting small errors, for an actual experiment several such fairly large models should be considered and compared to each other. There is no systematic way to find good models for this purpose, but any model with a large number of parameters can be used. For simplicity, we only regard one such model in this paper. This suffices for our purpose of determining how many measurements are needed as a function of both N and q.

The ideal GHZ state is permutationally invariant, in the sense that any permutation of its subsystems leaves the overall state unchanged. Mathematically, this can be expressed as

$$\rho_{\text{GHZ}} = \frac{1}{N!} \sum_{\pi_k \in S_N} V(\pi_k)\rho_{\text{GHZ}}V(\pi_k)^\dagger, \tag{6}$$

where the sum is over all $N!$ permutations π_k of the N qubits and $V(\pi_k)$ is the unitary representation of the operator that permutes the subsystems according to the permutation π_k. Since our simple three-parameter model is also permutationally invariant, it makes sense to use as the large model the set of *all* permutationally invariant (PI) states. This set has been

shown to be very convenient for quantum state reconstruction and entanglement detection [23–25]. Many experiments aim at generating GHZ states, *W* states, or Dicke states, all of which are PI. Moreover, if the PI part of a state ρ is entangled, then so is ρ.

(Note that this choice does not imply that we think the actual state is permutationally invariant, nor does it imply that we think the error process is permutationally invariant. All that matters is that our model will include the permutationally invariant part of the actual error process. As long as that part does not vanish, we will detect it. Recall that we *cannot* analyze all possible error models.)

As shown in Ref. [25], any permutationally invariant state can be represented as a block-diagonal matrix:

$$\rho_{\text{PI}} = \bigoplus_{j=j_{\min}}^{N/2} P_j \rho_j \otimes \frac{\mathbb{1}}{K_j}, \tag{7}$$

where ρ_j is a spin-j matrix of dimension $2j+1$, $\{P_j\}$ is a probability distribution over the spin values j, and K_j is the dimension of the non-PI part of the spin-j state, given by

$$K_j = \binom{N}{N/2 - j} - \binom{N}{N/2 - j - 1}. \tag{8}$$

The dimension of the permutationally invariant subspace grows as $\propto N^3$ with the number of qubits N. This model fits our purposes very well. For a dozen or more qubits the model contains a substantial number of parameters, but not so many that we cannot analyze it.

It was shown in Refs. [24,25] that the necessary and sufficient number of different measurements needed to gain full information about a permutationally invariant state is

$$D_N = \binom{N+2}{N}. \tag{9}$$

In particular, we can choose to measure observables of the form $\hat{A}^{\otimes N}$; that is, we can measure the same single-qubit observable on each qubit. We just have to pick D_N different single-qubit observables, the outcomes of which ought to be more or less uniformly distributed on the Bloch sphere [24,25].

III. NUMERICAL RESULTS

We simulate an experiment on N qubits. The "true" state that generates the data is chosen to be an unequal mixture of a noisy GHZ state ρ_{3P} (contained in the three-parameter model) and a randomly picked permutationally invariant state ρ_{PI} orthogonal to the three-parameter states (just to make sure the overlap of the actual state with the three-parameter subspace does not vary with N). We write

$$\rho_{\text{true}} = (1-q)\rho_{3P} + q\,\rho_{PI}. \tag{10}$$

The parameter q determines the probability of "wrong" types of errors, namely, those outside our standard error model. We simulate a certain number of measurements, where each single measurement consists of measuring N times the same single-qubit observable, where the latter is chosen from the set of D_N single-qubit observables. Thus, a single measurement yields N
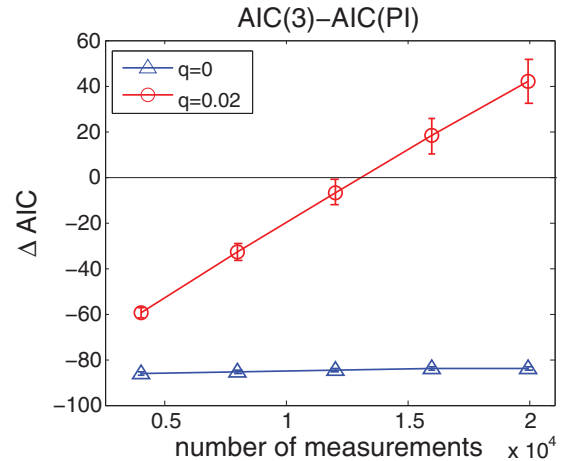

FIG. 1. (Color online) The differences in AIC values ΔAIC for a state of $N = 5$ qubits plotted against the total number of measurements. There are 21 measurements settings in this case, and the PI model contains 55 parameters. The simulation was run 100 times and the average ΔAIC is plotted. Error bars refer to the spread of ΔAIC over the 100 runs.

outcomes 0 or 1. We assume for simplicity that each of the D_N observables is measured the same number of times. We then find numerically the maximum likelihood state for the three-parameter model as well as for the large PI model. This is easy for the three-parameter model since the minimization is over just three parameters. For the PI model we apply an iterative algorithm described in [25] for which the required computation time increases only polynomially in the number of qubits. Using the two maximum likelihoods thus obtained, we can calculate the respective AIC values for the three-parameter and PI models and plot the difference, which we denote by ΔAIC. Negative values of ΔAIC correspond to the three-parameter model being favored, whereas positive values indicate that the PI model is better.

Figure 1 shows ΔAIC for two different "true" states, one with $q = 0$ and the other with $q = 0.02$. For $q = 0$ the Akaike information criterion correctly always prefers the three-parameter model. This is not as trivial (since the data are generated from a three-parameter state) as it may seem, because the statistical fluctuations are substantial (note that each observable is measured just a few dozen times for the smallest total number of measurements in the plot). For $q = 0.02$ we see that a relatively small number of measurements suffices to start favoring the PI model over the three-parameter model, and the more measurements one performs the firmer that conclusion gets. For very small numbers of measurements, a nonzero q cannot be detected yet, and we may interpret the point where ΔAIC crosses zero as the point where sufficiently many measurements have been taken to detect the presence of errors outside our standard (three-parameter) error model.

Let us consider how that crossing point changes with the number of qubits. A range of results for different N is plotted in Fig. 2(a). With increasing N the crossing point clearly moves toward larger numbers of measurements. We plot the crossing point as a function of N in Fig. 2(b). We see that the necessary total number of measurements to detect a fixed perturbation
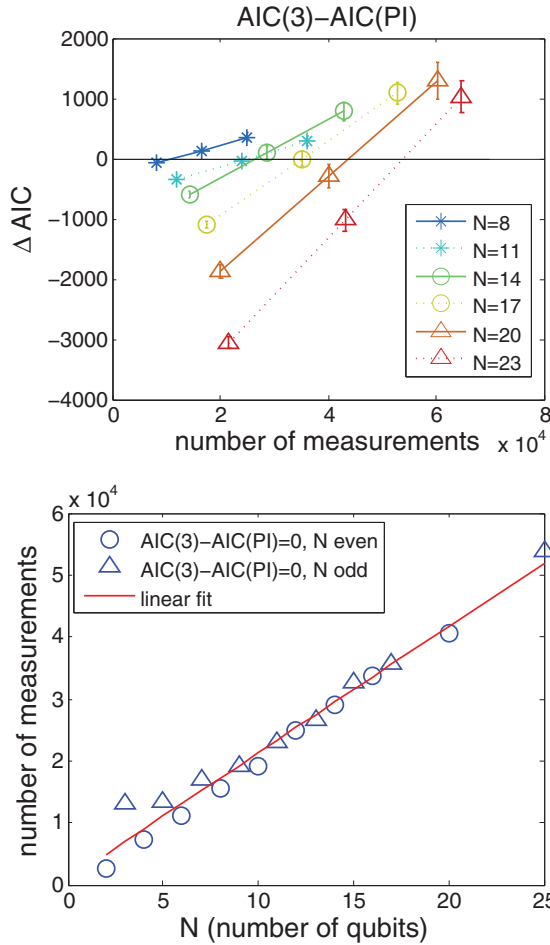
FIG. 2. (Color online) Plots for $q = 0.02$. (a) The differences in AIC for several different numbers of qubits, plotted against the total number of measurements. Note that a single measurement on N qubits yields N binary outcomes. For very small numbers of measurements, $\triangle$AIC approaches twice the difference in the number of parameters of the two models ($\approx N^3/3$). (b) The average number of measurements required to reach the point where both models are rated equally. Small even and odd numbers of qubits behave slightly differently.

FIG. 3. (Color online) (a) $\triangle$AIC as a function of the number of measurements performed, for four different values of q and $N = 5$ qubits. (b) The minimum number of measurements M needed to detect a perturbation of strength q, for both $N = 5$ and 10 qubits.

q increases only linearly in the number of qubits N, which shows that this can be measured very efficiently. (The number of single-qubit measurements needed grows as N^2.)

It is also useful to investigate how the number of measurements needed to detect a nonzero value of q depends on that value.

The plots of Fig. 3 show that the total number of measurements needed increases only moderately with $1/q$. This dependence becomes more favorable with increasing N, presumably because there are more ways to detect errors that occur with a given probability.

IV. CONCLUSIONS

We showed by example how to use the Akaike information criterion (AIC) to select between different error models in the context of quantum computing. Thanks to the AIC one does not need 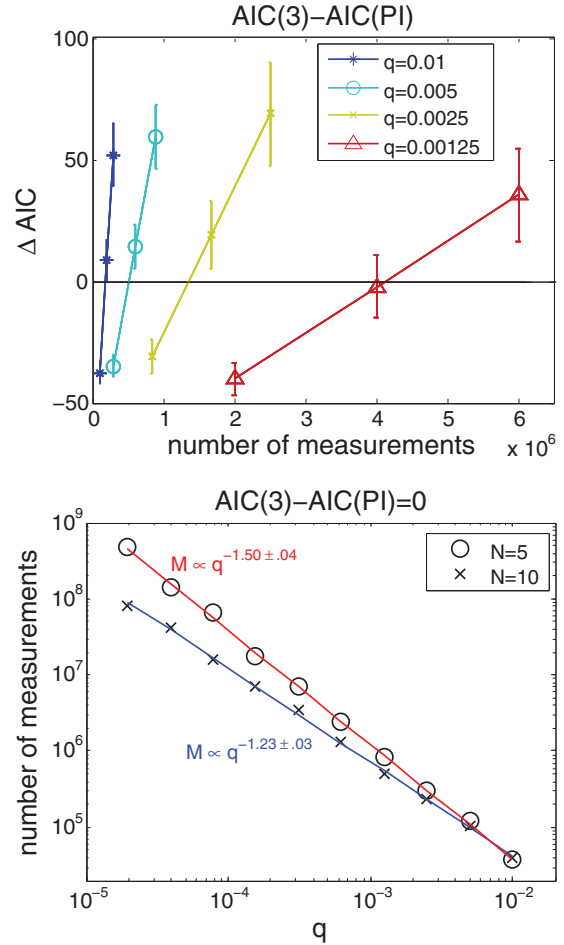exponentially many parameters to describe an experiment on multiple qubits. Instead, we compared a small model (with three parameters) with an intermediate-sized model [with $O(N^3)$ parameters]. The former stands for a standard error model in the context of fault tolerant quantum computing, and the larger model stands for other (undesired) types of errors. Since it is crucial to know whether one's implementation satisfies the condition for the fault tolerance error threshold theorems to apply, our method, which works for dozens of qubits, should be quite useful here. In our specific example the number of (unentangled) N-qubit measurements needed to detect errors of the wrong type turned out to scale linearly with the number of qubits and less than quadratically with the inverse of the wrong error probability. The latter scaling even improves with increasing number of qubits.

[1] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009).

[2] 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).

[3] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).

[4] T. Moroder, M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt, Phys. Rev. Lett. **110**, 180401 (2013).

[5] N. K. Langford, New J. Phys. **15**, 035003 (2013).

[6] S. J. van Enk and R. Blume-Kohout, New J. Phys. **15**, 025024 (2013).

[7] D. Gottesman, Phys. Rev. A **57**, 127 (1998).

[8] D. Aharonov and M. Ben-Or, *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (ACM, New York, 1997), pp. 176–188.

[9] E. Knill, R. Laflamme, and W. H. Zurek, Science **279**, 342 (1998).

[10] A. Y. Kitaev, *Quantum Communication, Computing, and Measurement* (Springer, New York, 1997), pp. 181–188.

[11] A. M. Steane, Phys. Rev. A **68**, 042322 (2003).

[12] A. G. Fowler, Phys. Rev. Lett. **109**, 180502 (2012).

[13] D. Gottesman, *Quantum Information Science and its Contributions to Mathematics*, Proceedings of Symposia in Applied Mathematics Vol. 68 (Amer. Math. Soc., Providence, 2009), p. 13.

[14] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging* (Cambridge University Press, Cambridge, 1993).

[15] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, New York, 2002).

[16] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **106**, 130506 (2011).

[17] H. Akaike, *Selected Papers of Hirotugu Akaike* (Springer, New York, 1998), pp. 199–213.

[18] K. Usami, Y. Nambu, Y. Tsuda, K. Matsumoto, and K. Nakamura, Phys. Rev. A **68**, 022314 (2003).

[19] P. Lougovski and S. J. van Enk, Phys. Rev. A **80**, 052324 (2009).

[20] J. O. S. Yin and S. J. van Enk, Phys. Rev. A **83**, 062110 (2011).

[21] M. Guţă, T. Kypraios, and I. Dryden, New J. Phys. **14**, 105002 (2012).

[22] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Am. J. Phys. **58**, 1131 (1990).

[23] G. Tóth, W. Wieczorek, R. Krischek, N. Kiesel, P. Michelberger, and H. Weinfurter, New J. Phys. **11**, 083002 (2009).

[24] G. Toth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010).

[25] T. Moroder, P. Hyllus, G. Toth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter, New J. Phys. **14**, 105001 (2012).