

Optimizing one-axis twists for variational Bayesian quantum metrology

Tyler G. Thurtell^{*} and Akimasa Miyake[†]

*Department of Physics and Astronomy, Center for Quantum Information and Control,
University of New Mexico, Albuquerque, New Mexico 87106, USA*

 (Received 13 February 2023; accepted 15 April 2024; published 16 May 2024)

Quantum metrology and sensing seek advantages in estimating an unknown parameter of some quantum state or channel, using entanglement such as spin squeezing produced by one-axis twists or other quantum resources. In particular, qubit phase estimation, or rotation sensing, appears as a ubiquitous problem with applications to electric field sensing, magnetometry, atomic clocks, and gyroscopes. By adopting the Bayesian formalism to the phase estimation problem to account for limited initial knowledge about the value of the phase, we formulate variational metrology and treat the state preparation (or encoding) and measurement (or decoding) procedures as parameterized quantum circuits. It is important to understand how effective various parametrized protocols are as well as how robust they are to the effects of complex noise such as spatially correlated noise. First, we propose a family of parametrized encoding and decoding protocols called arbitrary-axis twist ansatzes, and show that it can lead to a substantial reduction in the number of one-axis twists needed to achieve a target estimation error. Furthermore, we demonstrate that the estimation error associated with these strategies decreases with system size in a faster manner than classical (or no-twist) protocols, even in the less-explored regimes where the prior information is limited. Last, using a polynomial-sized tensor network algorithm, we numerically analyze practical variational metrology beyond the symmetric subspace of a collective spin and find that quantum advantage persists for the arbitrary-axis twist ansatzes with a few one-axis twists and smaller total twisting angles for practically relevant noise levels.

DOI: [10.1103/PhysRevResearch.6.023179](https://doi.org/10.1103/PhysRevResearch.6.023179)

I. INTRODUCTION

Sensors that utilize quantum mechanical effects to increase sensing precision and accuracy are expected to be among the first quantum information technologies to achieve broad applications [1]. The study of the abilities of such sensors is referred to as quantum sensing or quantum metrology. When the system being used as a sensor is a collection of qubits or spin-1/2 particles, a quintessential problem is qubit phase estimation, or rotation sensing [1,2]. Many applications can be understood as instances of this problem including magnetometry with neutral atoms [3], atomic clock stabilization with trapped ions [4] or with neutral atoms [5,6], nuclear magnetic resonance gyroscopes [7], Rydberg atom electric field sensors [8], and many other examples. This paradigm may even be relevant to quantum algorithms [9].

It has been known for a long time that using entangled many-body states can, in principle, lead to sensing precision that would not have been possible with an unentangled state of the same system [1,10–15]. Typically, the available advantage

is quadratic in system size, leading to the famous Heisenberg limit [12]. A useful type of entanglement for rotation sensing with spin-1/2 particles is spin squeezing, which can be generated, for example, via one-axis twisting-type interactions [16,17]. These interactions have been engineered in Bose-Einstein condensates [18], cavity QED setups [19], trapped ions [20], and approximately in Rydberg atom arrays [21,22]. When one-axis twisting is used in only the state preparation phase, the available advantage scales with system size to the power of 5/3 [17]. If one-axis twists are used in both the state preparation and measurement phases, as in twist-untwist protocols [23–30], the available advantage scales quadratically. These protocols have been realized experimentally, for example, in Ref. [29].

However, the situation is further complicated when prior knowledge about the value of the parameter to be estimated must be accounted for [31–45]. In addition, the extreme sensitivity of the required states to noise has so far prevented entanglement-enhanced sensors from becoming a practical reality [46–52]. A possible route to addressing these challenges is provided by the paradigm of hybrid quantum-classical variational techniques. In the context of quantum computing, variational quantum algorithms were introduced as an algorithmic framework that may be feasible even when the effects of noise cannot be suppressed to an arbitrarily low level [53–56]. More recently, variational approaches to quantum metrology have emerged as a potential application [34,57–73]. However, the most efficient use of available entangling resources is not known in most scenarios. Further, the

^{*}tthurtell@unm.edu

[†]amiyake@unm.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

robustness of these techniques to noise is not well understood. This second difficulty is partially due to the breakdown of standard symmetric subspace simulation techniques in the presence of spatially correlated noise. In this paper, we partially address both of these broad challenges.

First, to optimize the Bayesian mean-squared error (BMSE), we propose a family of parameterized circuits called arbitrary-axis twist ansatzes, where the entangling capacity of one-axis twists is effectively optimized with the help of global rotations about arbitrary axes. We find that performance comparable to that of previous approaches can be found with a reduced number of one-axis twists, especially in the few-twist regime. Second, we study whether these schemes remain effective in the presence of complex noise. To this end, we utilize a tensor network simulation technique which takes the permutation invariance of the noiseless evolution as a starting point to exactly simulate the protocol in the presence of correlated dephasing noise during the free evolution. We also study the effect of circuit-level noise associated with the one-axis twists.

The organization of the paper is as follows. In Sec. II, we review the necessary background. In Sec. III, we describe our arbitrary-axis twist ansatzes and evaluate their effectiveness. Finally, in Sec. IV, we study the robustness of our ansatzes to various types of noise.

II. BACKGROUND FOR QUANTUM METROLOGY

Throughout this paper, we consider a system of N qubits or spin- $\frac{1}{2}$ particles. We denote by X_j , Y_j , and Z_j , respectively, the Pauli- X , Y , and Z operators on qubit j . We will refer to the normalized sum of all Pauli operators in a particular direction, e.g., $J_z = \frac{1}{2} \sum_j Z_j$, as a component of the total angular momentum operator. We denote by $|0\rangle$ or $|1\rangle$, respectively, the $+1$ or -1 eigenstate of single qubit Pauli- Z and by $|+\rangle$ or $|-\rangle$, respectively, the $+1$ or -1 eigenstate of single qubit Pauli- X . We will also refer to any state of the form $e^{-i\theta_z J_z} e^{-i\theta_y J_y} |0\rangle^{\otimes N}$ for any azimuthal angle θ_z and any polar angle θ_y , as a spin coherent state.

A. Setting

The goal of single parameter quantum metrology [1,10–15,74,75] is to optimally estimate a parameter φ associated with some quantum channel Λ_φ . Since we are interested in optimal schemes for estimating the parameter, we must specify an objective function by which different schemes may be compared and a feasible set that describes schemes we are willing to consider. The exact form of the objective function is typically dictated by prior knowledge about the value the parameter may take. One method to properly account for prior knowledge (or lack thereof) about the value of φ is to treat φ as a random variable and pursue a Bayesian statistics approach to the problem. Now knowledge about φ before the experiment is performed can be captured by the prior distribution for this random variable $p(\varphi)$. A useful objective function is then given by the BMSE,

$$\Delta\varphi^2 \equiv \int_{-\infty}^{\infty} d\varphi \sum_{\zeta} [\varphi - \varphi_{\text{est}}(\zeta)]^2 p(\zeta|\varphi) p(\varphi), \quad (1)$$

where the $\{\zeta\}$ are measurement outcomes, $\varphi_{\text{est}}(\zeta)$ is an estimator for the value of φ , and $p(\zeta|\varphi)$ is the probability of measurement outcome ζ conditioned on the value of φ . This quantity is also referred to as the averaged estimation error. Such an approach is often called a global approach since it evaluates performance based on estimation error for all values of φ . In contexts where $\Lambda_\varphi[\cdot] = \Lambda_{\varphi+2\pi}[\cdot]$, as we will consider below, this integral is sometimes taken to run from 0 to 2π instead. Whether or not this choice makes sense depends on the physical origin of the channel $\Lambda_\varphi[\cdot]$. For example, a spin- $\frac{1}{2}$ particle exposed to a strong magnetic field for a fixed amount of time may acquire a relative phase between its spin-up and spin-down states that is equal to $2\pi n + \theta$ for some large integer n . This gives the same state as would a relative phase θ which may arise from a much weaker magnetic field but the underlying physical situation is clearly different. Further, when $\delta\varphi$ is not too large, the probability that $|\varphi| > 2\pi$ is small and the two approaches should give similar results. In this paper, the integrals in all BMSEs run from $-\infty$ to ∞ .

This is in contrast to approaches based on the quantum Fisher information [12,75] which tells us about the possible variance of the estimator at some particular parameter value φ_0 . To gain some understanding of the relationship between this averaged estimation error and the Fisher information, note that the BMSE can be written in terms of the variance and bias as

$$\Delta\varphi^2 = \int_{-\infty}^{\infty} d\varphi \{ \text{Var}_\varphi[\varphi_{\text{est}}] + \text{Bias}_\varphi[\varphi_{\text{est}}]^2 \} p(\varphi), \quad (2)$$

where the variance and bias are defined by

$$\text{Var}_\varphi[\varphi_{\text{est}}] \equiv \sum_{\zeta} (\varphi_{\text{est}}(\zeta) - \langle \varphi_{\text{est}} \rangle)^2 p(\zeta|\varphi), \quad (3)$$

$$\text{Bias}_\varphi[\varphi_{\text{est}}] \equiv \sum_{\zeta} (\varphi_{\text{est}}(\zeta) - \varphi) p(\zeta|\varphi), \quad (4)$$

and $\langle \cdot \rangle$ denotes the average over measurement outcomes. So, this roughly indicates that the Fisher information approach should be appropriate when an estimator can be found that is unbiased or has small bias for all possible values of parameter φ . Throughout this paper, we will take the BMSE as the objective function and we will take the prior to be a Gaussian distribution with mean zero and standard deviation $\delta\varphi$.

The basic structure of a quantum metrology experiment is outlined in Fig. 1. The first step is the preparation of some initial fiducial state ρ_{fid} . Throughout this paper, we take this state to be the spin-coherent state $\rho_{\text{fid}} = |+\rangle\langle+|^{\otimes N}$. An encoding channel, at least partially controlled by the experimenter, is then applied to this state to produce a probe state $\rho_{\text{probe}} = \Lambda_{\text{En}}[\rho_{\text{fid}}]$. By partially controlled, we mean, for example, that the intended probe encoding may consist of unitary dynamics, but the quantum channel that is actually realized may be a more general completely positive trace-preserving map due to experimental limitations. This situation is considered in Sec. IV. The probe state then undergoes a free evolution according to the channel $\Lambda_\varphi[\cdot]$ that is not controlled by the experimenter and depends on the parameter to be estimated. The resulting state now generally depends on the parameter $\varphi = \Lambda_\varphi[\rho_{\text{probe}}]$. We will consider channels that reduce to $\Lambda_\varphi[\cdot] = e^{-i\varphi J_z} \cdot e^{i\varphi J_z}$ in the limit that noise is negligible. After the encoding of the parameter φ , the experimenter can apply a

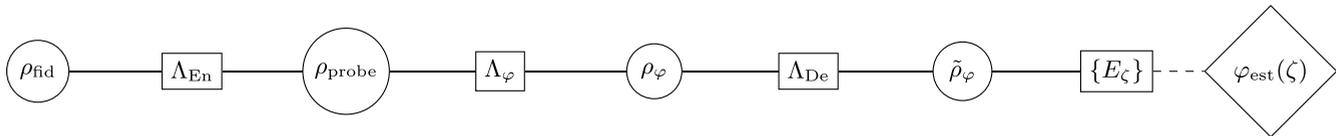


FIG. 1. A schematic representation of a general quantum metrology experiment. Time advances from left to right.

decoding channel $\Lambda_{\text{De}}[\cdot]$ to produce the state $\tilde{\rho}_\varphi = \Lambda_{\text{De}}[\rho_\varphi]$. At this point, the state is measured according to some positive operator-valued measure (POVM) with measurement operators $\{E_\zeta\}$ giving outcome ζ with probability $p(\zeta|\varphi) = \text{Tr}(\tilde{\rho}_\varphi E_\zeta^\dagger E_\zeta)$. We always take the measurement operators associated with this measurement to be projectors onto the space of states with Hamming weight w , i.e. $P_w = \sum_{x \text{ s.t. } |x|=w} |x\rangle\langle x|$ where $x \in \{0, 1\}^N$ and $|x|$ is the Hamming weight of x , i.e., the number of entries in the bit string x that are equal to one. This constitutes a measurement of the observable J_z . Note that these operators respect a permutation symmetry and can be implemented experimentally without the ability to detect each particle independently.

The final step in this type of experiment is to process the measurement outcome into an estimate $\varphi_{\text{est}}(w)$ of the phase φ . We take this estimator to be $\varphi_{\text{est}}(w) = \frac{a}{2}(N - 2w)$, where a is a proportionality constant. The average of this estimator over measurement outcomes is proportional to the expectation value of J_z in the state $\tilde{\rho}_\varphi$,

$$\begin{aligned} \langle \varphi_{\text{est}}(w) \rangle &\equiv \frac{a}{2} \sum_{w=0}^N p(w|\varphi)(N - 2w) \\ &= \frac{a}{2} \sum_{w=0}^N \text{Tr}(P_w \tilde{\rho}_\varphi)((N - w) + (-1)w) \\ &= a \langle J_z \rangle, \end{aligned} \quad (5)$$

where we have used the fact that

$$J_z = \frac{1}{2} \sum_w (N - w + (-1)w) P_w, \quad (6)$$

since each qubit in $|0\rangle$ increases the Z -angular momentum by $\frac{1}{2}$ and qubit in $|1\rangle$ decreases it by $\frac{1}{2}$. A permutation-invariant measurement operator such as this is known to be optimal in the absence of noise for this type of free evolution [31,33].

The optimal value of the proportionality constant a can be found by first expanding the quadratic form in Eq. (1) and performing the averages over the measurement outcomes and the prior distribution to obtain

$$\Delta\varphi^2 = \delta\varphi^2 - 2a(\varphi \langle J_z \rangle)_{\text{avg}} + a^2 \langle J_z^2 \rangle_{\text{avg}}, \quad (7)$$

where \cdot_{avg} denotes the average with respect to the prior $p(\varphi)$. Optimizing this expression with respect to a gives the optimal value of the proportionality constant

$$a_{\text{opt}} = \frac{(\varphi \langle J_z \rangle)_{\text{avg}}}{\langle J_z^2 \rangle_{\text{avg}}}. \quad (8)$$

This value is fixed by the encoding and decoding procedures and the free evolution. Notably, the value of a_{opt} depends on any noise that may be present. While the resulting estimator is not optimal over all estimators, it is simple to compute

and the use of a linear estimator eases comparison to prior work [34,69], where consideration was also restricted to linear estimators.

The optimal value of the BMSE over all possible probe states, measurements, and estimators can be found via an iterative optimization procedure [33,76,77]. In this optimization procedure, first the probe state is fixed and the measurement is optimized for that state. Then the measurement is fixed and the state is optimized for that measurement. This is then repeated until the procedure converges. Since in the absence of noise the optimal strategy is permutation invariant [31], this optimization can be restricted to the $N + 1$ dimensional permutation-invariant subspace. Unfortunately, however, the output of this optimization does not provide a procedure for realizing this optimal performance with any particular set of resources.

B. Practical resources for quantum metrology

We now turn the the question of what should constitute the feasible set of strategies. In many experimental setups, the available resources include global rotations

$$R_{\mathbf{n}}(\theta) = e^{-i\theta(\mathbf{n}\cdot\mathbf{J})} \quad (9)$$

and one-axis twists

$$T_{\mathbf{n}}(\theta) = e^{-i\theta(\mathbf{n}\cdot\mathbf{J})^2}, \quad (10)$$

where \mathbf{n} is a three-dimensional unit vector and $\mathbf{J} = J_x e_x + J_y e_y + J_z e_z$ [17,18,21,22]. Here e_μ is the unit vector directed along the μ axis. We will frequently use the shorthand notations $R_\mu(\theta) \equiv R_{e_\mu}(\theta)$ and $T_\mu(\theta) \equiv T_{e_\mu}(\theta)$ for global rotations and one-axis twists along the x , y , or z axes, respectively. Measurements of the total angular momentum of the ensemble along some axis are also commonly possible. Accordingly, we will take the feasible set of strategies to be defined by the set of probe states that can be prepared from $|+\rangle^{\otimes N}$ by a specified sequence of global rotations and one-axis twists and the set of POVMs that can be implemented via a (possibly different) specified sequence of global rotations and one-axis twists ending with a global $\frac{\pi}{2}$ x -rotation $R_x(\frac{\pi}{2})$ and followed by a measurement of the z component of the total angular momentum. Note that all of the entanglement used in these schemes comes from the one-axis twists. These are the resources used in twist-untwist protocols [23–30]. In these protocols, the encoding operation consists of a one-axis twist about the z axis by $\frac{1}{\sqrt{N}}$ followed by a rotation about the x axis by $\frac{\pi}{2}$. The decoding is then the inverse of the encoding operation followed by a final rotation about the x axis by $\frac{\pi}{2}$.

C. Variational metrology and parity symmetric ansatzes

The basic idea of quantum variational algorithms is to prepare, on the quantum device, a classically parameterized state by applying a series of unitaries that depend on classical

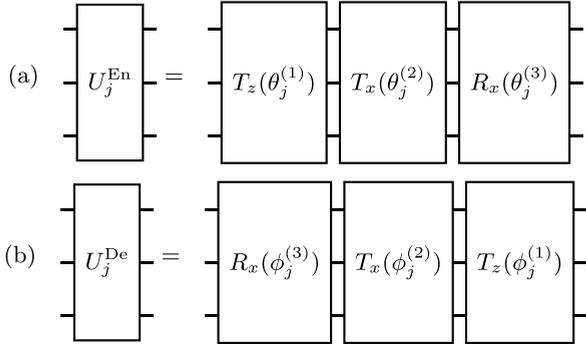


FIG. 2. Illustration of the general form of (a) encoding and (b) decoding layers used in the parity symmetric ansatzes [34,57]. In this diagram, operations further to the left occur first.

parameters to some fiducial quantum state following which some efficient measurement is performed on the resulting state. The outcome of this measurement, or a series of such measurements, is then used to update the values of the classical parameters. The sequence of unitaries used here is referred to as a parameterized circuit or an ansatz. We will replace both the encoding channel and the decoding channel described above with appropriate parameterized circuits and then numerically find the optimal values for the classical parameters.

We are primarily interested in studying the performance of two sets of parameterized circuits. The first set was introduced originally by the pioneering work of Kaubruegger *et al.* [34,57]. This family of ansatzes is described in terms of a number of encoding layers L_{En} and a number of decoding layers L_{De} , so the total encoding and decoding circuits can be written $U^{(\text{En/De})} = \prod_{j=1}^{L_{\text{En/De}}} U_j^{(\text{En/De})}$. Each layer is composed of two one-axis twists as defined in Eq. (10) and a global rotation as defined in Eq. (9). The encoding and decoding layers, respectively, have the form

$$U_j^{\text{En}} = R_x(\theta_j^{(3)})T_x(\theta_j^{(2)})T_z(\theta_j^{(1)}), \quad (11)$$

$$U_j^{\text{De}} = T_z(\phi_j^{(1)})T_x(\phi_j^{(2)})R_x(\phi_j^{(3)}), \quad (12)$$

as illustrated in Fig. 2. After the decoding unitary is applied, an $R_x(\frac{\pi}{2})$ rotation is performed, followed by a measurement of the z component of total angular momentum as described in Eq. (5). We call these the parity symmetric ansatzes since all gates used in the circuit independently commute with the X -parity operator $P_X = \prod_j X_j$. This property implies that the resulting estimator will be antisymmetric in φ and that the probe state's average angular momentum will be directed along the x axis. The second property is important because spin-squeezed states oriented along the x axis are likely to perform well. The member of this family of ansatzes with L_{En} encoding layers and L_{De} decoding layers will be referred to as the $\text{PAR}_{2L_{\text{De}}}^{2L_{\text{En}}}$ ansatz since $2L_{\text{En}}$ one-axis twists are used in the encoding and $2L_{\text{De}}$ twists are used in the decoding. The parity symmetric ansatzes were found to achieve nearly the optimal estimation performance allowed by quantum mechanics [33] when the number of layers is sufficiently large, see, for reference, Fig. 2 in Ref. [34].

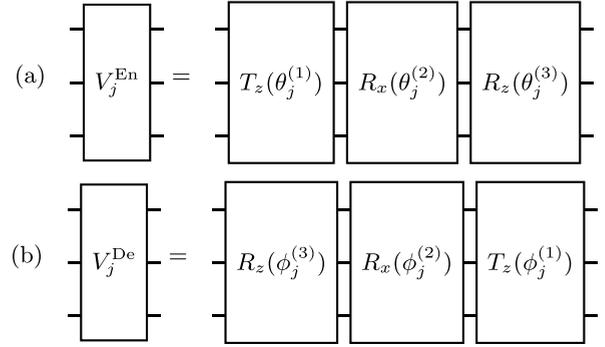


FIG. 3. Illustration of the general form of the unitaries that are appended to the (a) encoding and (b) decoding circuits to increase the number of one-axis twists once at least one is in use in the arbitrary-axis twist ansatzes. In this diagram, time proceeds from left to right.

III. ARBITRARY-AXIS TWIST ANSATZES

A. Definition and motivation

We propose a family of parameterized circuits for this problem, which we refer to as the arbitrary-axis twist ansatzes. The encoding and decoding unitaries are alternations of arbitrary global rotations and one-axis twists about the z axis. The unitaries always begin and end with an arbitrary rotation so they effectively implement a series of one-axis twists and global rotations about arbitrary axes followed by an additional rotation. We will label these ansatzes by the number of one-axis twists used in the encoding and the number of one-axis twists used in the decoding. The arbitrary-axis twist ansatz using n_{En} one-axis twists in the encoding and n_{De} one-axis twists in the decoding will be referred to as $\text{AAT}_{n_{\text{De}}}^{n_{\text{En}}}$. For example, AAT_1^1 can be thought of as a family of generalized twist-untwist protocols. The encoding and decoding unitaries used in AAT_1^1 are

$$V_1^{(\text{En})} = R_z(\theta_1^{(5)})R_x(\theta_1^{(4)})T_z(\theta_1^{(3)})R_z(\theta_1^{(2)})R_y(\theta_1^{(1)}), \quad (13)$$

$$V_1^{(\text{De})} = R_x(\phi_1^{(1)})R_z(\phi_1^{(2)})T_z(\phi_1^{(3)})R_x(\phi_1^{(4)})R_z(\phi_1^{(5)}). \quad (14)$$

As for the parity-symmetric ansatzes, we also include a final $R_x(\frac{\pi}{2})$ rotation at the end of the decoding immediately prior to the J_z measurement. Each additional one-axis twist requires three new parameters. In particular, adding an additional encoding or decoding one-axis twist means appending the following unitaries to the end (beginning) of the encoding (decoding) circuit:

$$V_j^{(\text{En})} = R_z(\theta_j^{(3)})R_x(\theta_j^{(2)})T_z(\theta_j^{(1)}), \quad (15)$$

$$V_j^{(\text{De})} = T_z(\phi_j^{(1)})R_x(\phi_j^{(2)})R_z(\phi_j^{(3)}). \quad (16)$$

The form of these appended unitaries is illustrated in Fig. 3. Altogether, $\text{AAT}_{n_{\text{De}}}^{n_{\text{En}}}$ contains $4 + 3(n_{\text{En}} + n_{\text{De}})$ circuit parameters.

We note that using sequential rotations about only two axes, rather than the standard three rotations by Euler angles, is sufficient to make the global rotations arbitrary. To understand why, notice that we can write a one-axis twist flanked

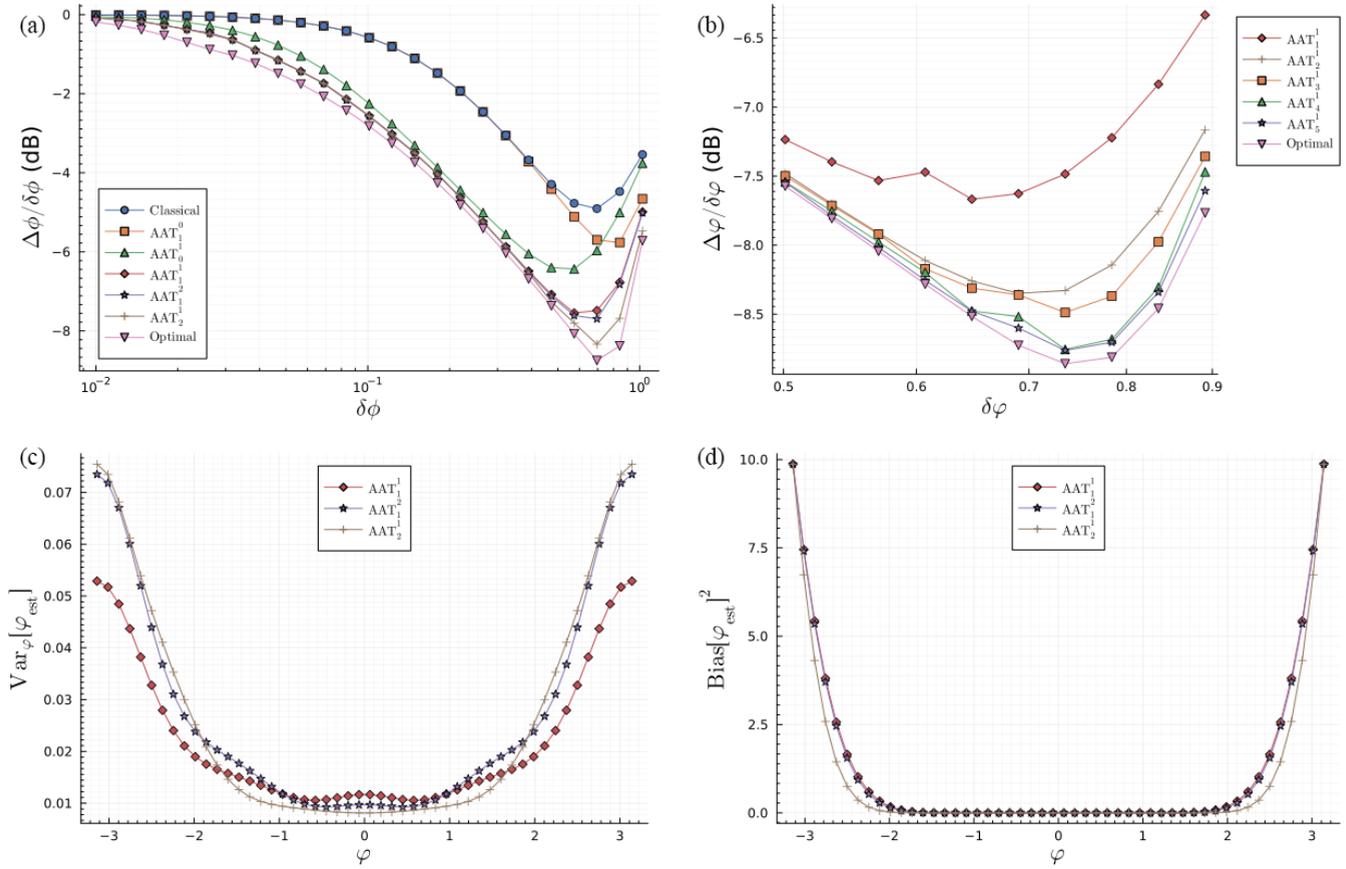


FIG. 4. (a) The optimized root Bayesian mean squared error $\Delta\phi/\delta\phi$ of various low depth arbitrary-axis twist ansatzes plotted vs the prior standard deviation $\delta\phi$. (b) Plot of $\Delta\phi/\delta\phi$ vs $\delta\phi$ for up to five decoding twists for $0.5 \lesssim \delta\phi \lesssim 0.9$. In (c) and (d), respectively, the variances and squared biases of the estimators associated with AAT₁¹, AAT₂¹, and AAT₂² are plotted vs φ at $\delta\phi \approx 0.74$. In all plots, $N = 30$.

by two different arbitrary global rotations as

$$\begin{aligned} R_z(\gamma')R_x(\beta')R_z(\alpha')T_z(\omega)R_z(\gamma)R_x(\beta)R_z(\alpha) \\ = R_z(\gamma')R_x(\beta')T_z(\omega)R_z(\alpha' + \gamma)R_x(\beta)R_z(\alpha). \end{aligned} \quad (17)$$

So, the number of rotations can be reduced by one. This can be repeated for each set of three rotations. Since the fiducial state is taken to be $|+\rangle$, a parameter can be removed from the first rotation by using the Euler decomposition $R_z(\gamma)R_y(\beta)R_x(\alpha)$ and dropping the x rotation. Similarly, since the final measurement is effectively a measurement of J_y , a parameter can be removed from the last global rotation by using the Euler decomposition $R_y(\gamma)R_x(\beta)R_z(\alpha)$ and dropping the last rotation.

Notably, while it utilizes only the one-axis twists about the fixed z axis, AAT _{n_{De}} ^{n_{En}} is the most general ansatz composed of one-axis twists and global rotations with n_{En} twists used in the encoding and n_{De} used in the decoding. This is because the rotations can effectively change the axis of the one-axis twists, so this family of ansatzes is universal to generate the set of unitaries which are composed of a sequence of one-axis twists about arbitrary axes interspersed with any global rotations. Due to the generality of the arbitrary-axis twist ansatzes, AAT _{n_{De}} ^{n_{En}} will perform at least as well as PAR _{n_{De}} ^{n_{En}} . However, PAR _{$2n_{\text{De}}$} ^{$2n_{\text{En}}$} has $3(n_{\text{En}} + n_{\text{De}})$ parameters while AAT _{$2n_{\text{De}}$} ^{$2n_{\text{En}}$} has $4 + 6(n_{\text{En}} + n_{\text{De}})$ parameters. Thus, comparisons between the members of these two families of ansatzes with a similar

number of parameters represent a comparison between the power of increased freedom in rotations and increased entanglement.

B. Noiseless performance

To analyze the performance of these ansatzes, we begin by numerically optimizing the circuit parameters for several of the low depth arbitrary-axis twist ansatzes over a wide range of values of the prior variance $\delta\phi$. While the simulation of the quantum evolution is efficient due to the permutation invariance in the ideal setting (and low-bond dimension matrix product state representations in the noisy setting), the classical optimization of the circuit parameters can be a major limitation as the number of circuit parameters is increased. Accordingly, we consider a system size of $N = 30$ throughout this section. See Appendix A for details on how the numerical optimization was performed. In Fig. 4(a), we plot the BMSE we obtain vs the prior standard deviation. In particular, we compare AAT₀¹, AAT₁¹, AAT₂¹, and AAT₂². For reference, we also include curves corresponding to the best entanglement-free (i.e., classical) approach and the optimal approach allowed by quantum mechanics [33]. We observe a first dramatic improvement in performance when moving from the classical approach to AAT₀¹. A second large improvement occurs when moving from AAT₀¹ to AAT₁¹. Finally, we notice another significant improvement when moving from

AAT₁¹ to AAT₂¹. This indicates that, once there is a single encoding twist, the most promising place to look for improvement is in adding twists to the decoding. This also makes sense in light of previous observations that the optimal probe states are well approximated by spin squeezed states in this case [33].

We now focus on the regime where the largest improvement over the classical approach is possible, i.e., $0.5 \lesssim \delta\varphi \lesssim 0.9$. In this region, we examine how increasing the depth of the decoding circuit affects the BMSE. A comparison for AAT₁¹, AAT₃¹, AAT₄¹, and AAT₅¹ is shown in Fig. 4(b). We observe that as the circuit depth is increased, the optimized estimators approach the optimal value of the BMSE. However, we also note that increasing the circuit depth appears to provide diminishing returns as the optimal is approached. While this is to some extent inevitable due to the existence of a lower bound, we observe strikingly little improvement in the transition from AAT₄¹ to AAT₅¹. Due to these diminishing returns, it is reasonable to consider circuits with only a few one-axis twists in at least some cases.

We next examine the origin of the improvement of AAT₂¹ over AAT₁¹. To do this, we plot, in Fig. 4(c), the variance of the optimized estimators at $\delta\varphi \approx 0.74$ over a range of values of φ for AAT₁¹, AAT₁², and AAT₂¹. We note that neither AAT₂¹ nor AAT₁² exhibit dramatic improvement over AAT₁¹. In Fig. 4(d), we plot the bias squared of the optimized estimators over measurement outcomes at the same value of $\delta\varphi$ and same range of φ for AAT₁¹, AAT₁², and AAT₂¹. We observe that the estimator associated with AAT₂¹ displays a larger region of low bias than the other two ansatzes. This is in line with previous observations that increasing the depth of the decoding circuit primarily serves to decrease bias [69]. It is also not surprising that a decrease in bias should be the most effective way to obtain a smaller BMSE since we are examining the case where the prior variance is fairly large.

C. System size dependence and scaling

Next, we examine how the performance of our ansatz compares to that of the optimal and classical strategies as the system size is increased with an emphasis on scaling. The quantum van Trees [34,42,78–80], or Bayesian quantum Cramér-Rao, inequality states that

$$\Delta\varphi^2 \geq \frac{1}{F_Q + F_P}, \quad (18)$$

where F_Q is the quantum Fisher information averaged over the prior distribution and F_P is the Fisher information of the prior distribution. In the absence of noise, the quantum Fisher information is independent of the value of φ and thus, this averaged Fisher information satisfies

$$F_Q \leq 4(\text{Tr}(J_z^2 \rho_{\text{probe}}) - \text{Tr}(J_z \rho_{\text{probe}})^2) \leq N^2. \quad (19)$$

The Fisher information of a Gaussian distribution is given by

$$F_P = \frac{1}{\delta\varphi^2}. \quad (20)$$

Motivated by this, we simplify the curve fitting process by considering a shifted version of the BMSE defined by

$$(\Delta\varphi')^{-2} \equiv \Delta\varphi^{-2} - \delta\varphi^{-2}. \quad (21)$$

The fitting is further simplified by subtracting out the excess mean-squared error associated with phase slips, i.e., times when the $\phi < \pi$ or $\phi > \pi$. When this happens the mean-squared error is approximately $4\pi^2$. Further, the probability of this occurring is $1 - \text{erf}(\frac{\pi}{\sqrt{2}\delta\varphi})$. Accordingly, we further transform to

$$\Delta\tilde{\varphi}^2 \equiv \Delta\varphi'^2 - 4\pi^2 \left(1 - \text{erf}\left(\frac{\pi}{\sqrt{2}\delta\varphi}\right)\right). \quad (22)$$

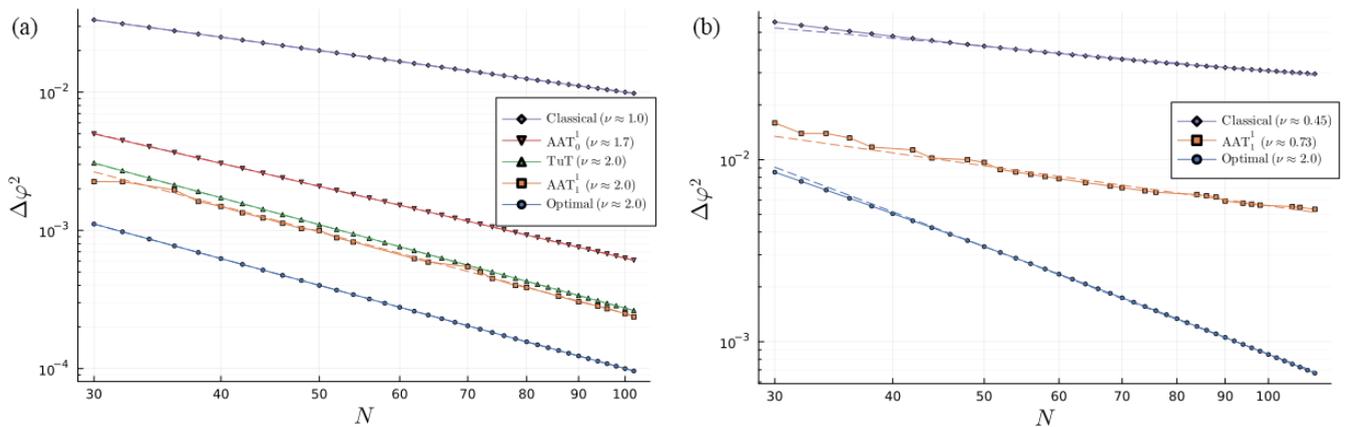


FIG. 5. Plots of the shifted Bayesian mean squared error $\Delta\tilde{\varphi}^2$ vs system size N for prior standard deviations of (a) $\delta\varphi = 10^{-3}$ and (b) $\delta\varphi = 0.7$. The markers on the solid curves correspond to the performance of the strategies the dashed lines are the fitted curves. In (a), the optimal, AAT₂¹, twist-untwist (TuT) strategies are all fit to curves with exponents $\nu \approx 2.0$, the AAT₀¹ is fit to a curve with $\nu \approx 1.7$, and the classical strategy is fit to a curve with $\nu \approx 1.0$. In (b), the optimal strategy is again fit to a curve with $\nu \approx 2.0$, the AAT₁¹ strategy is fit to a curve with $\nu \approx 0.73$, and the classical strategy is fit to a curve with $\nu \approx 0.45$.

To study the scaling, we then fit $\Delta\tilde{\varphi}$ to a curve of the form

$$f(N) = \frac{\alpha}{N^\nu}, \quad (23)$$

where α and ν are fitting parameters. When performing the fitting, we exclude points where optimization becomes stuck in a local minimum.

The results are shown in Fig. 5 for prior standard deviations of (a) $\delta\varphi = 10^{-3}$ and (b) $\delta\varphi = 0.7$. The markers on the solid line indicate the performance of the strategy at a particular system size and the dashed lines indicate the fittings. In Fig. 5(a), where the prior standard deviation is small, the BMSE is dominated by variance and the optimal and classical curves respectively exhibit the familiar Heisenberg $\nu \approx 2$ and standard quantum limit $\nu \approx 1$ scaling. Our AAT₀¹ shows a scaling in good agreement with that expected of spin squeezed states $\nu \approx 1.7 \approx \frac{5}{3}$ [17]. Our AAT₁¹ ansatz exhibits Heisenberg scaling for large enough system sizes. For reference, we also plot the performance of the twist-untwist (TuT) protocol described in Ref. [25]. The twist-untwist protocol exhibits comparable but slightly worse performance than our AAT₁¹ ansatz which makes sense given that the AAT₁¹ ansatzes are generalized twist-untwist protocols. In Fig. 5(b), where the prior standard deviation is much larger, we find that the exponents of the optimal strategy continues to exhibit Heisenberg scaling as expected from the π -corrected Heisenberg limit [38]. However, while the optimal scaling is independent of the prior standard deviation, the variational strategies only achieve the optimal performance at large prior standard deviation if a large number of layers are used. For this reason, we see a decrease in the scaling exponent when the prior standard deviation is increased but the number of layers is kept small. The scaling of the classical strategy decreases to $\nu \approx 0.45$. This decrease may be due to the restriction that $\langle \varphi_{\text{est}} \rangle \propto \langle J_z \rangle$. Our AAT₁¹ ansatz exhibits an intermediate scaling with $\nu \approx 0.73$. From these observations, we infer that even our low depth arbitrary-axis twist ansatzes can achieve better scaling than the corresponding classical strategies over a wide range of values of prior standard deviation.

D. Comparison to the previously considered parity symmetric ansatzes

Finally, we compare the performance of the optimized arbitrary-axis twist estimators to that of the optimized parity symmetric estimators with the low- and medium-depth decoding circuits introduced in Refs. [34,57,69]. The results of this comparison are shown in Fig. 6. We find that AAT₁¹ achieves a smaller BMSE than PAR₂² despite using only half as many one-axis twists. We also observe that AAT₄¹ performs comparably to PAR₆² despite the former using only five one-axis twists and the latter using eight. We also have some evidence that the optimization of the arbitrary-axis twist ansatzes seems to be eased for numerical optimization. In particular, we find that arbitrary-axis twist ansatzes may be less likely to fail due to local minima compared to the parity symmetric ansatzes with a similar number of parameters. See Appendix C for more details about the comparison of the performance of these two families of ansatzes and Appendix B for information about the total twisting times associated with the arbitrary-axis twist strategies.

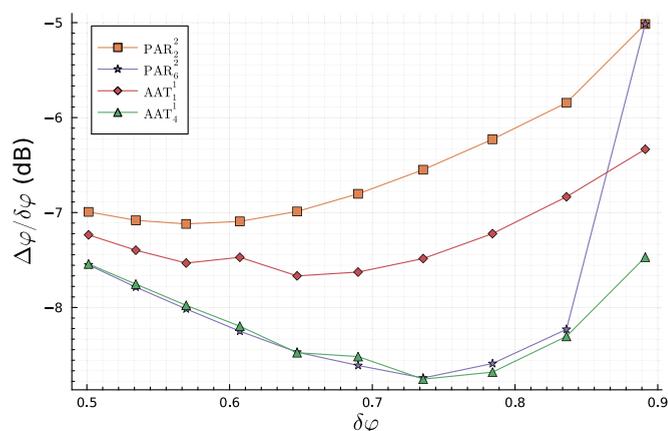


FIG. 6. Plot of the root Bayesian mean squared error $\Delta\varphi/\delta\varphi$ vs prior variance $\delta\varphi$ for the optimized ansatzes PAR₂², PAR₆², AAT₁¹, and AAT₄¹ in the $0.5 \lesssim \delta\varphi \lesssim 0.9$ regime for $N = 30$. AAT₁¹ performs better than PAR₂² and AAT₄¹ performs about as well as PAR₆². We note that the rightmost PAR₆² point is trapped in a local minima in this numerical setting, but can be resolved by a different initialization, for instance. See Appendixes A and C for details on how the optimization was performed and values of the Bayesian mean squared error that can be obtained with this ansatz and prior variance via other optimization methods.

IV. TENSOR NETWORK SIMULATION FOR NOISY QUANTUM METROLOGY

As the effect of single qubit dephasing noise during the free evolution in variational quantum metrology has been studied previously in both the Bayesian [34] and Fisher information settings [59,62,64,65,71,73], here we focus on the effect of more complex noise such as correlated dephasing models. We also study the effect of circuit level noise during the encoding and decoding circuits.

A. Correlated dephasing

In realistic scenarios, the signal φ to be detected may not be constant over all qubits. We consider the case where the signal φ_j at each site can be broken up as $\varphi_j = \varphi + r_j$, where φ is a constant piece at each site and r_j is a Gaussian random variable with mean zero. In particular, we take the spatial correlation functions associated with this random variable to be $\bar{r}_j r_k \equiv C_{j,k}$, where $\bar{\cdot}$ indicates the average over the Gaussian distribution that the r_j are drawn from and we will assume, as a first nontrivial example, that $C_{j,k} = 0$ if $|j - k| > 1$. The physical content of this is that we allow for correlations but only between nearest-neighbor sites in 1D. Such an evolution can be thought of as a global z rotation followed by a correlated dephasing channel. The effective noise channel is then

$$\begin{aligned} \mathcal{N}[\rho] &= \overline{e^{-i\sum_j \frac{r_j}{2} Z_j} \rho e^{i\sum_j \frac{r_j}{2} Z_j}} \\ &= \sum_{m,n \in \{0,1\}^N} \rho_{m,n} \overline{e^{-\frac{i}{2}(\sum_{j,k} (-1)^{m_j} r_j - (-1)^{n_k} r_k)}} |m\rangle \langle n| \\ &= \sum_{m,n \in \{0,1\}^N} \rho_{m,n} e^{-\frac{1}{8} \sum_{j,k} ((-1)^{m_j} - (-1)^{n_j}) C_{j,k} ((-1)^{m_k} - (-1)^{n_k})} |m\rangle \langle n|, \end{aligned} \quad (24)$$

where in the last line we explicitly performed the multivariate Gaussian integration. Such a noise model may be relevant to cavity QED-based experiments in which the spins are arranged in a quasi-one-dimensional lattice [19,29,30,81,82].

Up until this point, all dynamics have been restricted to the symmetric subspace, i.e., the $\frac{N}{2}(\frac{N}{2} + 1)$ eigenspace of $J_x^2 + J_y^2 + J_z^2$. This subspace has dimension $N + 1$ allowing for efficient simulation. This is a result of the initial states, unitary dynamics, and measurements respecting a permutation symmetry. However, this noise model breaks the permutation symmetry. It has been shown by Chabuda *et al.* in Ref. [77] that this noise model nevertheless has an efficient tensor network representation, and a tensor network contraction was utilized to solve recursively for the optimal Fisher information associated with this noisy free evolution. On the other hand, we propose compiling symmetric subspace techniques to tensor network constructions for permutation-invariant states and one-axis twists and combine them with the above construction of correlated dephasing, so the correlated noise effects during the free evolution of metrology protocols are simulated efficiently.

B. Matrix product states and operators for permutation-invariant states and one-axis twists

We now review some basic facts about matrix product states (MPSs) and operators (MPOs) [83–86] and efficient representations of permutation-invariant states and one-axis twists. An MPS representation of a state $|\psi\rangle$ is a representation of the form

$$|\psi\rangle = \sum_{x \in \{0,1\}^N} \prod_j A^{(x_j)}[j]|x\rangle, \quad (25)$$

where the $\{A^{(0)}[j], A^{(1)}[j]\}$ are matrices, except when $j = 1$ or $j = N$, in which case they are, respectively, row and column vectors. The label x_j is called the physical index. The largest dimension of any of the matrices $A^{(s)}[j]$ is referred to as the bond dimension of the representation. Any state of N qubits can be put into an MPS form with a bond dimension of $2^{\lceil \frac{N}{2} \rceil}$.

A natural extension of the MPS construction is the MPO. An MPO representation of an operator \mathcal{O} is a representation of the form

$$\mathcal{O} = \sum_{x^L, x^R \in \{0,1\}^N} \prod_j A^{(x_j^L, x_j^R)}[j]|x^L\rangle\langle x^R|. \quad (26)$$

The labels x_j^L and x_j^R are called physical indices and the matrix indices are referred to as virtual indices. All operators on N qubits can be put into the form of an MPO with bond dimension $4^{\lceil \frac{N}{2} \rceil}$.

An MPS representation for the state that results from acting on an operator with a known MPO representation on a state with a known MPS representation can be determined by contracting the physical indices of the MPO associated with x^R with the physical indices of the MPS. If the initial state MPS representation had bond dimension d_s with tensors composed of $\{A^{(x_j)}[j]\}$ and the MPO had bond dimension d_o with tensors composed of $\{B^{(x_k^L, x_k^R)}[j]\}$, then the MPS representation for the

final state has bond dimension $d_s d_o$ with

$$A^{(x_j)}[j] = \sum_s B^{(x_j, s)}[j] \otimes A^{(s)}[j]. \quad (27)$$

This allows for the simulation of time evolution by contracting an MPO representation of the time evolution operator with an MPS representation of some initial state. If the bond dimension of the state remains polynomial at all times, then expectation values of operators that have polynomial bond dimension MPO representations can be computed with polynomial cost.

An MPS representation with bond dimension $\lceil \frac{N}{2} \rceil + 1$ can always be constructed for a permutation-invariant state as follows. Any permutation-invariant state of N qubits can be written as

$$|\psi\rangle = \sum_{x \in \{0,1\}^N} c_{|x|}|x\rangle, \quad (28)$$

where the coefficients $c_{|x|}$ depend only on the Hamming weight of x . The matrices used in the matrix product state representation at a site $j < \lceil \frac{N+1}{2} \rceil$ have dimension $j \times (j + 1)$ and elements given by

$$(A_S^{(0)}[j])_{m,n} = \delta_{m,n}, \quad (29)$$

$$(A_S^{(1)}[j])_{m,n} = \delta_{m,n-1}. \quad (30)$$

The matrices associated with site $N - j + 1$ with $j > \lceil \frac{N+1}{2} \rceil$ are just the transpose of those associated with site j . The remaining site $j = \lceil \frac{N+1}{2} \rceil$ has matrices

$$(A_S^{(0)}[\lceil (N+1)/2 \rceil])_{m,n} = c_{m+n}, \quad (31)$$

$$(A_S^{(1)}[\lceil (N+1)/2 \rceil])_{m,n} = c_{m+n+1}. \quad (32)$$

Analogously, all permutation-invariant operators have a MPO representation with bond dimension $O(N^3)$ as described in Appendix D.

We now turn to constructing MPOs for one-axis twists. We focus on the one-axis twist about the z axis but similar results hold for any operator that is permutation-invariant and can be diagonalized by single qubit operations. A one-axis twist about the z axis can be written as

$$T_z(\theta) = \sum_{x \in \{0,1\}^N} e^{-\frac{i}{4}(N-2|x|)^2} |x\rangle\langle x|. \quad (33)$$

The key observation is that, as was the case for the permutation-invariant state, the coefficients depend only on the Hamming weight. The MPO matrices for a site $j \neq \lceil \frac{N+1}{2} \rceil$ are then given by $A^{(0,0)}[j] = A_S^{(0)}[j]$, $A^{(1,1)}[j] = A_S^{(1)}[j]$, and $A^{(0,1)}[j] = A^{(1,0)}[j] = 0$. At site $\lceil \frac{N+1}{2} \rceil$, the matrices are given by

$$(A^{(0,0)}[\lceil (N+1)/2 \rceil])_{m,n} = e^{-\frac{i}{4}(N-2(m+n))^2}, \quad (34)$$

$$(A^{(1,1)}[\lceil (N+1)/2 \rceil])_{m,n} = e^{-\frac{i}{4}(N-2(m+n+1))^2}, \quad (35)$$

and $A^{(0,1)}[\lceil (N+1)/2 \rceil] = A^{(1,0)}[\lceil (N+1)/2 \rceil] = 0$.

The other unitary element of our ansatzes, the global rotations, can be given a bond-dimension 1 representation. For example, the matrices at each site associated with an x rotation

are

$$A^{(\mu,\nu)}[j] = \langle \mu | e^{-i\frac{\theta}{2}X} | \nu \rangle, \quad (36)$$

where $\mu, \nu \in \{0, 1\}$. Accordingly, when used in time evolution, the global rotations do not cause the bond dimension of the state to increase. This is a result of the fact that the global rotations do not involve interactions between spins.

Finally, we will need MPO constructions for the operators J_z and J_z^2 since we must evaluate expectation values of these operators. The operator J_z can be given a bond-dimension 2 representation. At sites away from the boundary $j \neq 1, N$, the matrices are given by

$$A_{J_z}^{(\mu,\nu)}[j] = \begin{pmatrix} 1 & (-1)^{\mu\frac{1}{2}} \\ 0 & 1 \end{pmatrix} \delta_{\mu,\nu}. \quad (37)$$

At $j = 1$ and $j = N$, the matrices are, respectively, $A_{J_z}^{(\mu,\nu)}[1] = \begin{pmatrix} 1 & (-1)^{\mu\frac{1}{2}} \\ 0 & 1 \end{pmatrix} \delta_{\mu,\nu}$ and $A_{J_z}^{(\mu,\nu)}[N] = \begin{pmatrix} (-1)^{\mu\frac{1}{2}} \\ 1 \end{pmatrix} \delta_{\mu,\nu}$. The operator J_z^2 can be given a bond-dimension 3 representation. At sites away from the boundary, it has matrices given by

$$A_{J_z^2}^{(\mu,\nu)}[j] = \begin{pmatrix} 1 & (-1)^{\mu\frac{1}{2}} & \frac{1}{4} \\ 0 & 1 & (-1)^{\mu} \\ 0 & 0 & 1 \end{pmatrix} \delta_{\mu,\nu}. \quad (38)$$

At the boundary sites, the matrices are given by $A_{J_z^2}^{(\mu,\nu)}[1] = \begin{pmatrix} 1 & (-1)^{\mu\frac{1}{2}} & \frac{1}{4} \\ 0 & 1 & (-1)^{\mu} \\ 0 & 0 & 1 \end{pmatrix} \delta_{\mu,\nu}$ and $A_{J_z^2}^{(\mu,\nu)}[N] = \begin{pmatrix} (-1)^{\mu\frac{1}{2}} \\ (-1)^{\mu} \\ 1 \end{pmatrix} \delta_{\mu,\nu}$.

C. Simulation algorithm

We use these constructions as a basis for simulations of metrology experiments in the presence of noise that breaks the permutation symmetry. If 1D geometrically local noise, for example, the correlated dephasing described above, occurs during the free evolution, we introduce a simulation strategy that makes use of both symmetric subspace simulation and matrix product constructions.

Initially, the first three stages in Fig. 1 are simulated in the symmetric subspace so an $(N+1) \times (N+1)$ symmetric subspace representation of ρ_{probe} is obtained. Then this is converted to an MPO representation using the construction for permutation-invariant states described in the last section.

Second, if the only entangling resources used are one-axis twists and the number of one-axis twists is sufficiently small, the decoding, i.e., stages 6 and 7 in Fig. 1, can be dealt with in two ways. The first method is to map each one-axis to an MPO as described in the last section. The MPOs for all one-axis twists and global rotations used in the decoding can then be contracted to obtain an MPO for the time evolution operator associated with the decoding. Alternatively, if the number of one-axis twists is large enough that the first procedure would lead to an intractable bond dimension or entangling resources other than one-axis twists are used, the decoding unitary may be compiled into a single permutation-invariant unitary as described in Appendix E. If the depth of the circuit is ℓ , then the cost of this compilation is $O(\ell N^7)$ in time. Note that it is not sufficient to work only in the symmetric subspace when compiling the decoding unitary, as it will be acting on a state

that lies outside of the symmetric subspace. The resulting permutation-invariant unitary can then be mapped to an MPO with bond dimension $O(N^3)$ via a construction analogous to the one for permutation-invariant states as noted above and in Appendix D. The first approach has the advantage of allowing the compiling step to be skipped, while the advantage of the second approach is that by construction the bond dimension of the resulting MPO is guaranteed to be polynomial in system size. In this paper, we utilize the former approach in which the MPOs for individual one-axis twists are independently constructed.

Next, an MPO is constructed for the superoperator associated with the free evolution, i.e., stages 4 and 5 in Fig. 1. The correlated dephasing model described above in Eq. (24) has an MPO representation with bond-dimension 2 [77]. Since this noise commutes with the unitary part of the free evolution we consider and the unitary part is a global rotation, the entire free evolution superoperator can be given a bond-dimension 3 representation.

Finally, the MPOs obtained in these three steps can then be contracted to obtain an MPO of the state immediately before measurement. If the number of one-axis twists used in the decoding does not scale with system size and the MPO for the free evolution superoperator has polynomial bond dimension, either decoding simulation strategy gives an algorithm with cost polynomial in system size. If the number of gates used in the decoding scales polynomially with system size then the compiling strategy continues to give a polynomial cost algorithm.

D. Robustness to correlated noise

We now study the robustness of the strategies found to be optimal in Sec. III to various types of noise. First, in the presence of spatially correlated dephasing as described in Sec. IV A, we analyze how the performance of the solutions found to be optimal in the absence of noise decays. In particular, we take the correlation functions to be

$$\overline{r_j r_k} = c_1 \delta_{j,k} + c_2 (\delta_{j,k-1} + \delta_{j,k+1}), \quad (39)$$

so c_1 is the variance at each site and c_2 is the covariance between adjacent sites. These results were obtained using our tensor network simulation scheme. Due to the polynomial bond dimensions of the involved states and operators described in Sec. IV B, there is no truncation of singular values and the results are exact to numerical precision.

The effect of this type of noise on the optimal AAT₁¹ noiseless strategy found at $\delta\varphi \approx 0.74$ for different values of c_1 and c_2 is shown in Fig. 7. Note that while we use the circuit parameters from the noiseless optimization, the proportionality constant a in the estimator is changed to be the optimal one given the noise channel defined by Eq. (24). We do not observe a dramatic drop off in performance in the presence of correlated noise. This is indicated by the relatively constant value of the BMSE along curves of constant $c_1 + c_2$ when $c_2 > 0$. We also find that the BMSE at constant values of c_1 decreases as c_2 is made more negative. In other words, performance is improved as the noise becomes anti-correlated. This can be understood as resulting from the

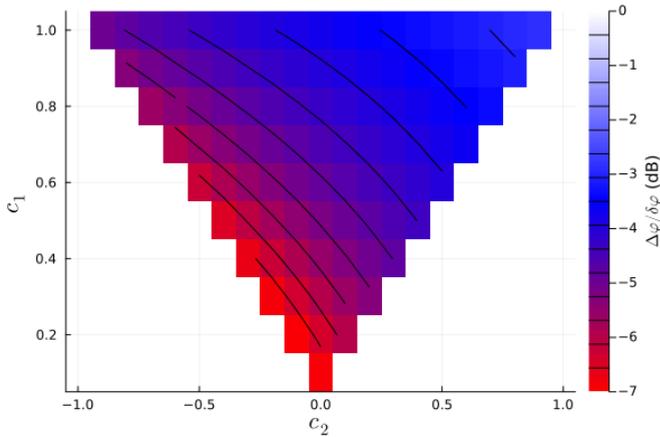


FIG. 7. Reduction of the root Bayesian mean-squared error $\Delta\varphi/\delta\varphi$ for the optimized AAT_1^1 ansatz under a correlated dephasing channel of Eq. (24) during a free evolution when $N = 30$ and $\delta\varphi \approx 0.74$. The correlated noise is characterized in terms of the variance of r_j at each site c_1 and the covariances at adjacent sites c_2 . The curves of constant $\Delta\varphi/\delta\varphi$ indicate that correlated dephasing is not much worse than uncorrelated, and anticorrelated dephasing is more favorable than uncorrelated.

underlying estimator being proportional to $\langle J_z \rangle$ when averaged over measurement outcomes. Intuitively, as the estimator is the average of a separate estimator for each qubit, we may imagine some cancellation of the errors if the errors in these estimators are anti-correlated.

E. Circuit level noise

Finally, we study the robustness of the strategies found to be optimal in Sec. III to circuit level noise, namely, noises during encoding and decoding circuits. First, we consider the effect of single-qubit dephasing,

$$\mathcal{D}_p^{(j)}[\rho] = (1 - p)\rho + pZ_j\rho Z_j, \quad (40)$$

on each qubit after each one-axis twist, as a model of a noisy one-axis twist. We refer to p as the noise strength. We again employ the tensor network based simulation and again the proportionality constant a used in the estimator is the optimal one given the noise. The results are shown in Fig. 8(a). We observe that AAT_1^1 and AAT_2^1 outperform the classical strategy up to a noise strength of at least $p = 0.1$. Second, notice that AAT_2^1 performs better than AAT_1^1 over the same range. This is not a property shared by the ansatzes with deeper decoding circuits. The deeper circuit ansatzes are all outperformed by AAT_2^1 starting at some noise strength

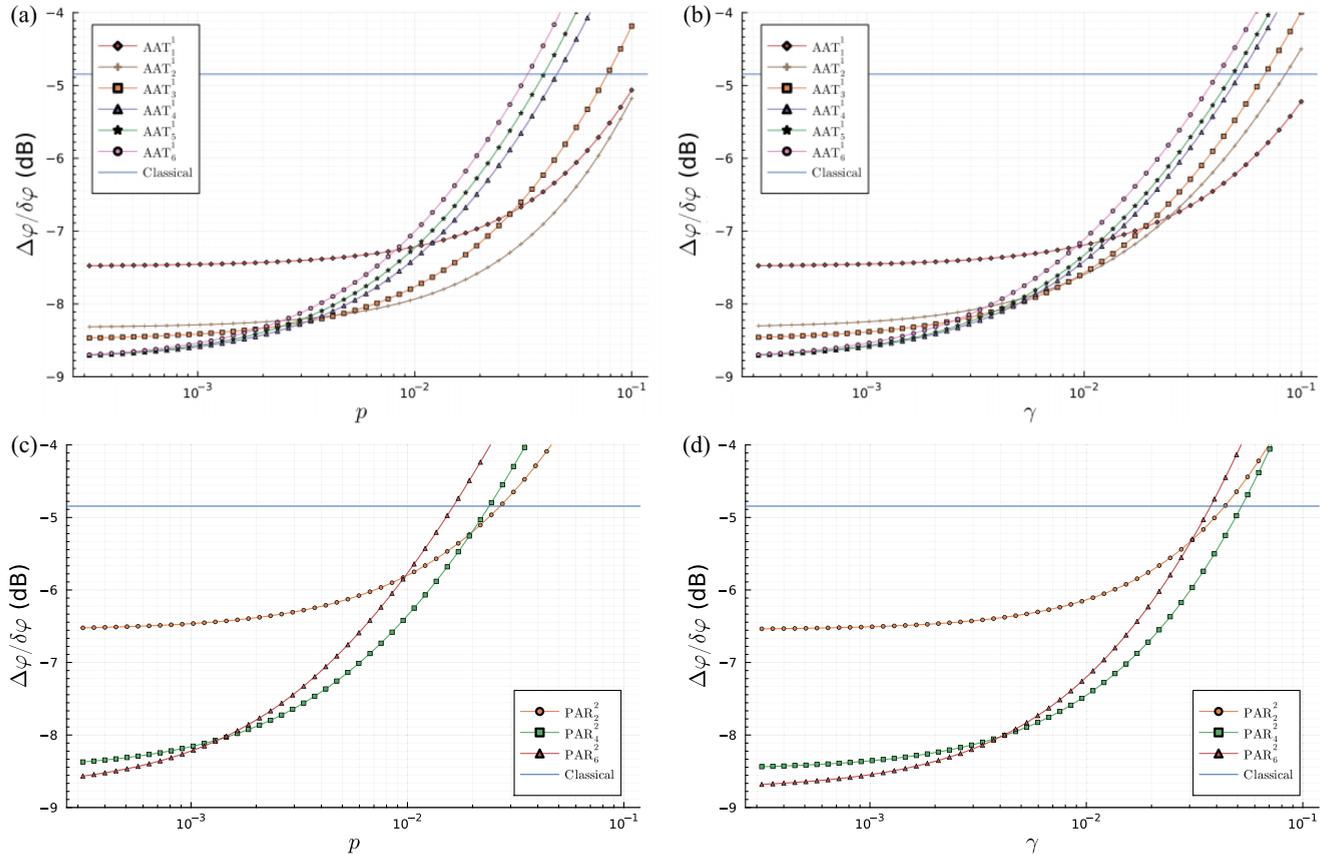


FIG. 8. The root Bayesian mean squared error $\Delta\varphi/\delta\varphi$ for (a), (b) AAT_1^1 , AAT_2^1 , AAT_3^1 , AAT_4^1 , AAT_5^1 , AAT_6^1 ; (c), (d) PAR_2^2 , PAR_4^2 , and PAR_6^2 , and the classically optimal strategy under noisy one-axis twists vs the noise strength when $N = 30$ and $\delta\varphi \approx 0.74$. In (a) and (c), the noise model is dephasing noise with noise strength p and in (b) and (d) the noise model is amplitude damping noise with noise strength γ . The ansatzes with deeper decoding circuits are advantageous only when the noise strength p or γ is sufficiently small.

between 0.0025 and 0.0047. AAT_3^1 is outperformed by AAT_1^1 at $p \approx 0.03$ and by the classical strategy at $p \approx 0.079$. The ansatzes with deeper decoding circuits are outperformed by AAT_1^1 beginning between a noise strength of $p \approx 0.0085$ and $p \approx 0.012$ and they are outperformed by the classical strategy beginning between a noise strength between $p \approx 0.035$ and $p \approx 0.05$. In Fig. 8(c), the robustness of the parity symmetric ansatzes to dephasing noise is displayed. The behavior is qualitatively similar. We note that AAT_1^1 always outperforms PAR_2^2 , AAT_4^1 always outperforms PAR_4^2 and PAR_6^2 , and AAT_2^1 outperforms PAR_4^2 and PAR_6^2 for noise strengths greater than about 0.00091. In particular, for noise strengths greater than 0.00091, all of the considered parity symmetric ansatzes are outperformed by at least one of AAT_1^1 or AAT_2^1 . All of the parity symmetric strategies are outperformed by the classical strategy for error rates greater than $p \approx 0.027$ due to the increased number of one-axis twists.

Next, we consider the effect of single qubit amplitude damping. This could arise due to atomic spontaneous emission. The noise channel is given by

$$\begin{aligned} \mathcal{A}_\gamma^{(j)}[\rho] = & \frac{(1 + \sqrt{1 - \gamma})^2}{4} \rho + \frac{(1 - \sqrt{1 - \gamma})^2}{4} Z_j \rho Z_j \\ & + \frac{\gamma}{4} \{Z_j, \rho\} + \gamma \sigma_j^{(-)} \rho \sigma_j^{(+)}, \end{aligned} \quad (41)$$

where γ is the noise strength for this model and $\sigma_j^{(\pm)} = \frac{1}{2}(X_j \mp iY_j)$. Again we consider the situation in which each qubit passes through this channel after each one-axis twist. As for the last noise model, the estimator coefficient used is the optimal one given the noise. The results are shown in Fig. 8(b). For this noise model, AAT_1^1 still outperforms the classically optimal strategy up to at least $\gamma = 0.1$. All deeper ansatzes are outperformed by AAT_1^1 at some noise strength between 0.0095 and 0.028 and by the classical strategy at some noise strength between 0.044 and 0.089. The robustness of the optimized parity symmetric strategies to the same noise is shown in Fig. 8(d). Again, the results are qualitatively similar but we note that AAT_1^1 always outperforms PAR_2^2 and AAT_2^1 outperforms PAR_4^2 and PAR_6^2 for noise strengths greater than 0.0047. Above this noise strength, all parity-symmetric ansatzes are outperformed by at least one of AAT_1^1 or AAT_2^1 . All of the parity-symmetric strategies are outperformed by the entanglement-free strategy for values of γ greater than ≈ 0.0056 .

The first point of this analysis is to give some indication of when it is worth increasing the circuit depth in a particular device, as which circuit performs the best can be dependent on the noise level. The second is to indicate that the presence of noise during the encoding and decoding can rapidly eliminate the benefit of increased decoding depth. This is another sense in which going to deeper ansatzes provides diminishing returns.

V. CONCLUSION

We have advanced several aspects of Bayesian, or global, variational quantum metrology. We introduced and analyzed a family of parameterized circuits which we call arbitrary-axis twist ansatzes. These are the most general ansatzes composed of one-axis twists and global rotations with a fixed number

of one-axis twists in the encoding and decoding circuits. We found that a reduction in the number of one-axis twists needed for performance comparable to that of a previously considered variational approach based on parity symmetric ansatzes is available. This is especially meaningful in the shallow depth regime where the reduction can be by as much as half. The significance of this finding is complemented by another finding that the usefulness of deeper circuits is harmed by the presence of noise in the encoding and decoding circuits. In the example shown in Fig. 8, we found that for dephasing noise strengths greater than $p = 0.00091$ or amplitude damping noise strengths greater than $\gamma = 0.0047$, all parity symmetric ansatzes considered are outperformed by AAT_2^1 . At larger noise strengths, they are also all outperformed by AAT_1^1 . In addition to being the ansatzes utilizing the fewest one-axis twists, AAT_1^1 and AAT_2^1 are usually also among the ansatzes with the smallest total twisting angles (see Appendix B). In fact, for the example in Fig. 8 they are the two ansatzes with the smallest total twisting angle. They also start outperforming the other AAT ansatzes at similar noise strengths. Minimizing the number of entangling resources is also useful from a classical simulation perspective. For example, we simulated the effect of spatially correlated dephasing noise during the free evolution using a tensor network algorithm. The cost of this algorithm was reduced due to the small number of entangling one-axis twist gates in the ansatz of interest.

Our findings first show the rather important contributions of nonentangling resources to variational metrology especially in the shallow depth regime. This could be a useful fact in the design of future ansatzes. Second, we found that variational metrology can still perform well in the presence of a small amount of spatially correlated noise. Finally, we also studied the robustness of the ideal strategies to circuit level noise, and found that for deep decoding circuits the performance falls off rapidly with the strength of the noise associated with the one-axis twists. This suggests that in the absence of some form of quantum error correction, variational metrology circuits should be kept shallow.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation QLCI Q-SEnSE Grant No. OMA-2016244, and STAQ Project No. PHY-1818914. We thank the UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high performance computing resources used in this work.

APPENDIX A: NUMERICAL DETAILS

We numerically search for the optimal values of the circuit parameters for a given value of $\delta\varphi$ for both the arbitrary-axis twist and parity symmetric ansatzes. We first perform the optimization for AAT_1^1 and PAR_2^2 . Then the circuit parameters found via this optimization are used as the initialization for one level deeper circuits, e.g., AAT_2^1 . The parameters found via this optimization are then used as the starting point for the next level deeper optimization and so on. We refer to this initialization strategy as sequential initialization.

The optimization of each ansatz begins with a Nelder-Mead search [87–89]. For an optimization over n variables, Nelder-Mead constructs an $n + 1$ point simplex. Roughly

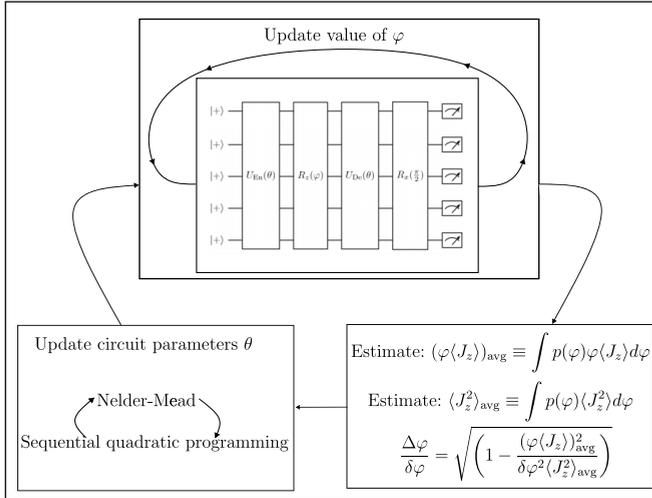


FIG. 9. Illustration of the algorithm used to find the optimal strategies. First, the expectation values of J_z and J_z^2 are estimated at a sufficient number of different values of φ to estimate $\Delta\varphi$. Then the estimate of $\Delta\varphi/\delta\varphi$ and possibly estimates of the derivatives of this function are used to choose new values of the circuit parameters θ to consider.

speaking, at each step of the optimization, it attempts to replace the point in the simplex with the largest objective value with a better point. If a better point cannot be found, this is interpreted as indicating the presence of a valley and the simplex is shrunk. The optimization is stopped when the difference between the largest and smallest value currently in the simplex differ by less than some value ϵ_1 .

We then take the output parameters from the Nelder-Mead search and use them as the initialization parameters for a sequential quadratic programming optimization [90,91]. At each iteration, this algorithm finds the minimum of a quadratic approximation to the objective function at the current point. While this approach can provide better performance than linear approximations, it does require access to second derivatives. The stopping criteria is that the minimum objective function value between consecutive iterations differ by less than ϵ_2 . We then take the output circuit parameters from that optimization and use them to initialize an additional Nelder-Mead optimization. This process is repeated until consecutive optimizations agree to within ϵ_3 .

While the values of ϵ_1 , ϵ_2 , and ϵ_3 could, in principle, be allowed to differ, we fix them all to the same value $\epsilon = \epsilon_1 = \epsilon_2 = \epsilon_3$ and take that value to be $\epsilon = 10^{-13}$. The optimization is carried out using the software package NLOPT [92] with automatic differentiation provided by ZYGOTE [93]. We also note that the numerical integration over φ to estimate the BMSE is performed using Hermite-Gauss quadrature with 500 integration points during the optimization and 25 integration points in the tensor network simulations. The 25 integration points are close to the amount used in a recent proof of principle experiment [69]. The total algorithm is illustrated by the flow chart in Fig. 9. The Hermite-Gauss quadrature is performed using the FASTGAUSSQUADRATURE package [94] and the tensor network simulations use the ITENSORS package [95,96].

TABLE I. This table contains the values of $\theta_{\text{En}}^{\text{twist}}$, $\theta_{\text{De}}^{\text{twist}}$, and $\theta_{\text{tot}}^{\text{twist}}$ for $n_{\text{En}} = 1$, $N = 30$, $\delta\varphi \approx 0.74$, and various values of n_{De} . The values reported here are for three significant figures.

$(n_{\text{En}}, n_{\text{De}})$	$\theta_{\text{En}}^{\text{twist}}$	$\theta_{\text{De}}^{\text{twist}}$	$\theta_{\text{tot}}^{\text{twist}}$
(1,1)	0.0600	0.0331	0.0931
(1,2)	0.0630	0.0713	0.134
(1,3)	0.0642	0.172	0.236
(1,4)	0.0675	0.186	0.253
(1,5)	0.0680	0.194	0.262
(1,6)	0.0685	0.262	0.331

APPENDIX B: RESOURCE COUNTING AS MEASURED BY THE NUMBER OF ONE-AXIS TWISTS VS TOTAL TWISTING ANGLE

To place in context the method of resource counting utilized in this paper, we briefly comment on the total twisting angle $\theta_{\text{tot}}^{\text{twist}}$, which is given by the sum of the magnitudes of the angles associated with all one-axis twists used in the ansatz, as a potential alternative to the number of one-axis twists. Although the total twisting angle is a seemingly simple way to quantify the cost of the variational strategies for many experimental setups, in this Appendix we will explain why the number of one-axis twists is indeed both a more suitable and more tractable resource to count. We also note that the strategies considered here do not exhibit large total twisting angles in practice, in fact they are usually less than $\frac{\pi}{8}$.

First, in Table I, we separately display the total twisting angle associated with the encoding circuit $\theta_{\text{En}}^{\text{twist}}$ and the total twisting angle associated with the decoding circuit $\theta_{\text{De}}^{\text{twist}}$ for $N = 30$ and $\delta\varphi \approx 0.74$ (our most commonly studied prior standard deviation in the main text). The data suggests that these should be treated as two separate objects rather than two parts of $\theta_{\text{tot}}^{\text{twist}}$. For example, when the data from this table is fit to a curve of the form

$$f(n_{\text{En/De}}) = An_{\text{En/De}} + B, \quad (\text{B1})$$

$\theta_{\text{En}}^{\text{twist}}$ exhibits almost constant behavior with $A \approx 0.0017$ and $B \approx 0.059$, while $\theta_{\text{De}}^{\text{twist}}$ exhibits a stronger dependence on n_{De} with $A \approx 0.044$ and $B \approx 0.00041$. This is closely related to the observation in the main text that most of the improvement comes from adding twists to the decoding once a single twist is used in the encoding. In Fig. 10, we plot the total twisting angle vs decoding depth for the optimized AAT ansatzes with $N = 30$, $n_{\text{En}} = 1$, and various values of $\delta\varphi$. Usually, AAT ansatzes with a larger number of twists result in a larger total twisting angle than AAT ansatzes with fewer twists. In the cases where this is not true, we often observe little improvement due to the additional twists. These observations suggest that the number of one-axis twists used by the strategy is indeed a reasonable proxy for the total twisting angle for the AAT ansatzes.

Second, the optimization where the number of twists used in the encoding and in the decoding are separately allowed to vary but the total twisting angle is fixed is possible but could lead to unusual solutions if the number of twists is truly unbounded. In particular, such an approach could result

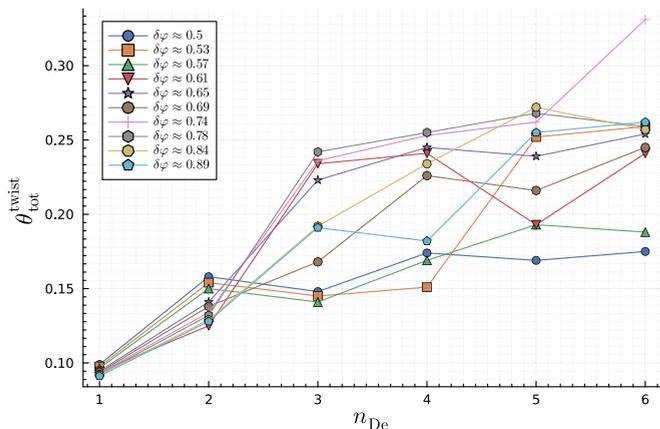


FIG. 10. Plot of the total twisting angle of the optimal $\text{AAT}_{n_{\text{De}}}^1$ strategies vs the decoding depth for various values of $\delta\varphi$. Note that there is a clear relation between the two quantities and in the case that $\delta\varphi \approx 0.74$ the relationship is nearly linear for the depths considered here.

in pathological solutions where the number of twists used becomes very large but the angle of each twist becomes very small. Such a strategy may be challenging to implement in some experimental platforms, for example, in ion traps [69]. Additionally, it is still guaranteed that at a fixed total twisting angle the AAT ansatzes will outperform the PAR ansatzes, as the latter are a special case of the former. We also note that fixing the number of twists does provide an upper bound on the total twisting angle used for that optimization. In this sense, the number of twists can be viewed as a coarse-grained measure to capture much of the physics of the dependence of the BMSE on the magnitude of the twisting angles. For completeness, we mention the total twisting angle for the PAR ansatzes for $\delta\varphi \approx 0.74$. The total twisting angles associated with PAR_2^2 and PAR_4^2 are both about 0.153, and the total angle associated with PAR_6^2 is about 0.194. These total twisting

angles are larger than that of AAT_2^1 but smaller than that of AAT_4^1 . However, the total twisting angle varies for other values of $\delta\varphi$. In particular, the total twisting angle of AAT_4^1 has a range between 0.15 and 0.26 as seen in Fig. 10, so it can be comparable to that of PAR_6^2 . In particular, for some values of $\delta\varphi$, AAT_4^1 exhibits a smaller total twisting angle than PAR_6^2 .

APPENDIX C: COMPARISON OF ARBITRARY-AXIS TWIST AND PARITY SYMMETRIC ANSATZES

We note that the result of the optimization, and even which ansatzes outperform others, can depend significantly on the specifics of the hyperparameters used. To study this phenomena, we perform the optimization of PAR_6^2 and AAT_3^1 for different sets of hyperparameters. In particular, the hyperparameters we adjust are ϵ , whether the sequential initialization is used or if all circuit parameters are initialized to zero, and the number of points used to approximate the integral over φ . The results of all these optimizations are shown in Fig. 11. We chose these two ansatzes as examples for three reasons. First, they are fairly deep circuits. Second, PAR_6^2 and AAT_3^1 have a similar number of circuit parameters with 12 and 16, respectively. Third, they are an example where which ansatz achieves a smaller BMSE depends on the hyperparameters used. We find that the optimizations are liable to get stuck in local minima whenever the hyperparameters are relaxed from the values described above.

This sensitivity to hyperparameters has the potential to be an issue for experimental on device optimization since it necessitates examining more values of φ for the integration and using more repetitions to increase the estimation precision of $\Delta\varphi/\delta\varphi$. We also note that we have found that the sequential initialization strategy described above is able to at least partially mitigate these issues. We have also heuristically observed that the arbitrary-axis twist ansatzes may be somewhat less susceptible to these issues than the parity symmetric ansatzes. On an optimistic note, most optimizations

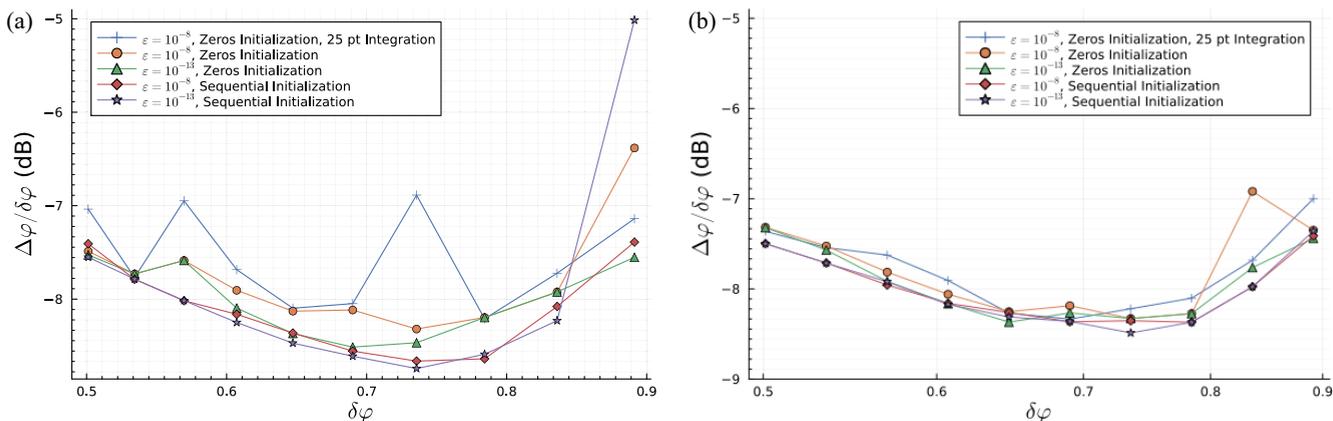


FIG. 11. Plot of the root Bayesian mean-squared error $\Delta\varphi/\delta\varphi$ vs prior variance $\delta\varphi$ resulting from the Nelder-Mead sequential quadratic programming approach described in the text for various hyperparameters for (a) PAR_6^2 and (b) AAT_3^1 . The blue crosses are the results for initializing all circuit parameters to zero, setting $\epsilon = 10^{-8}$, and using only 25 points in the numerical integration. The orange circles are the same as the blue crosses but with 500 points used in the numerical integration. The green triangles are the same as the orange circles but with $\epsilon = 10^{-13}$. The red diamonds use $\epsilon = 10^{-8}$ but sequential initialization. Finally, the purple triangles are the result of using the hyperparameters described in Appendix A. As they are usually the best, the numerical setting of the purple triangles are displayed throughout the main text

found reasonably good values for $\Delta\varphi/\delta\varphi$ even if they do not find the global minimum. This is likely acceptable for some applications.

APPENDIX D: MPOs FOR PERMUTATION INVARIANT OPERATORS

In this Appendix, we provide a construction of an MPO for any permutation invariant operator with bond dimension $O(N^3)$. The action of the permutation group defines a set of equivalence classes of computational basis states labeled by the Hamming weight according to

$$|x\rangle \sim |y\rangle \quad \text{iff } \exists \sigma \in S_N \text{ s.t. } x = \sigma(y). \quad (\text{D1})$$

For reasons that will become clear shortly, in this Appendix and the next we will denote the Hamming weight as t_1 . It is useful to choose a representative element from each of these equivalence classes. We choose elements of the form $|x(\mathbf{t})\rangle = |0\rangle^{\otimes N-t_1} \otimes |1\rangle^{\otimes t_1}$ so the first $N - t_1$ qubits are in state $|0\rangle$ and the rest are in state $|1\rangle$. A permutation invariant state $|\psi\rangle$ is then one of the form

$$|\psi\rangle = \sum_{\substack{\mathbf{t} \\ \text{s.t. } t_1 \leq N}} \frac{c_{\mathbf{t}}}{t_1!(N-t_1)!} \sum_{\sigma \in S_N} P_{\sigma} |x(\mathbf{t})\rangle, \quad (\text{D2})$$

where the denominator of the coefficient is a normalization convention which accounts for repetitions of identical kets and $P_{\sigma} |x\rangle \equiv |\sigma(x)\rangle$.

This framework can be naturally extended to states of N qudits. We denote a standard basis for qudits of dimension d by $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$. A standard basis state for N qudits of dimension d then has the form $\otimes_j |s_j\rangle \equiv |s_1 s_2 \dots s_N\rangle$, where each $|s_j\rangle \in \{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$. Now the equivalence classes defined by the action of the permutation group are labeled by a type vector of the form $\mathbf{t} = (t_1, t_2, \dots, t_{d-1})$ subject to the constraint $\sum_{j=1}^{d-1} t_j \leq N$. We again choose a representative element from each equivalence class as $|x(\mathbf{t})\rangle = |0\rangle^{\otimes N - \sum_j t_j} \otimes |1\rangle^{\otimes t_1} \otimes \dots \otimes |d-1\rangle^{\otimes t_{d-1}}$. It is convenient to define $t_0 \equiv N - \sum_j t_j$ to be the number of qudits in state $|0\rangle$. Then permutation invariant qudit states $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{\substack{\mathbf{t} \\ \text{s.t. } \sum_j t_j \leq N}} \frac{c_{\mathbf{t}}}{t_0! \prod_j t_j!} \sum_{\sigma \in S_N} P_{\sigma} |x(\mathbf{t})\rangle. \quad (\text{D3})$$

The number of independent coefficients here is equal to the number of distinct type vectors or the number of multisets of cardinality N with elements drawn from an underlying set of cardinality d , i.e., $\binom{N+d-1}{d-1} = O(N^{d-1})$.

This framework can also include permutation invariant operators via the Choi-Jamiołkowski isomorphism. Under this isomorphism, an operator $\mathcal{O} = \sum_{s,s'} c_{s,s'} |s\rangle\langle s'|$ on qudits of dimension d is associated with a state of qudits of dimension d^2 according to

$$\begin{aligned} \mathcal{O} &= \sum_{s,s'} c_{s,s'} |s\rangle\langle s'| \\ \rightarrow |\mathcal{O}\rangle &= \sum_{s,s'} c_{s,s'} |(s_1, s'_1)(s_2, s'_2) \dots (s_N, s'_N)\rangle, \end{aligned} \quad (\text{D4})$$

where the standard d^2 dimensional qudit states are now labeled by the ordered pairs (s_j, s'_j) and the s_j can take d

values. To make the consequences for the underlying operator transparent, when considering an operator on qubits we will write the type vectors, labeling the equivalence classes as $\mathbf{t} = (t_{(0,1)}, t_{(1,0)}, t_{(1,1)})$ and identify a representative from each equivalence class as $|x(\mathbf{t})\rangle = |(0,0)\rangle^{\otimes N-t_{(0,1)}-t_{(1,0)}-t_{(1,1)}} \otimes |(0,1)\rangle^{\otimes t_{(0,1)}} \otimes |(1,0)\rangle^{\otimes t_{(1,0)}} \otimes |(1,1)\rangle^{\otimes t_{(1,1)}}$. This representative element can be connected via the Choi-Jamiołkowski isomorphism to an operator on qubits:

$$\begin{aligned} \hat{x}(\mathbf{t}) &= |0\rangle\langle 0|^{\otimes N-t_{(0,1)}-t_{(1,0)}-t_{(1,1)}} \\ &\times \otimes |0\rangle\langle 1|^{\otimes t_{(0,1)}} \otimes |1\rangle\langle 0|^{\otimes t_{(1,0)}} \otimes |1\rangle\langle 1|^{\otimes t_{(1,1)}}. \end{aligned} \quad (\text{D5})$$

A permutation invariant operator is then one that can be written as

$$\mathcal{O} = \sum_{\substack{\mathbf{t} \\ \text{s.t. } \sum_j t_j \leq N}} \frac{c_{\mathbf{t}}}{t_{(0,0)}! \prod_j t_j!} \sum_{\sigma \in S_N} P_{\sigma} \hat{x}(\mathbf{t}) P_{\sigma}^{\dagger}, \quad (\text{D6})$$

where, in analogy with the qudit case, $t_{(0,0)} \equiv N - \sum_j t_j$.

It has been known for a long time that permutation invariant states of qubits can be represented with a bond dimension that scales linearly with system size [83]. Here we give a concrete construction that works for any state of qudits of any dimension d in the symmetric subspace and has maximum bond dimension $O(N^{d-1})$. That is, the resulting MPS has a bond dimension that is polynomial in the system size. This construction can be extended to MPOs for operators via the Choi-Jamiołkowski isomorphism.

In this construction, the tensor at site 1 is described by $1 \times d$ matrices of the form $(A^{(s)}[1])_{1,m} = \delta_{s,m-1}$. This vector can be thought of as keeping track of the type vector of some given basis ket in the sense that there is one possible vector with a single entry equal to one and the rest zero here for each possible type vector of a one-site system. The remaining tensors associated with sites on the left half of the chain, i.e., those sites with $j < \lceil \frac{N+1}{2} \rceil$, will serve to update this vector with new information about the global type as we move from left to right in that $\prod_{k=1}^j A^{(s_k)}[k]$ should be a $1 \times \binom{j+d-1}{d-1}$ vector in which each entry is associated with one of the possible type vectors of a j site system. Accordingly, the right dimension of $A^{(s)}[j]$ is $\binom{j+d-1}{d-1} = O(j^{d-1})$. The sorting of the entries of this vector is in order of increasing values of $\|\mathbf{t}\|_1 = \sum_j t_j$. For types with the same value of $\|\mathbf{t}\|_1$, the entries are sorted first in order of decreasing t_1 . For types with the same value of t_1 , they are sorted in order of decreasing t_2 and so on.

If $s = 0$, then $(A^{(0)}[j])_{\mu,\nu} = \delta_{\mu,\nu}$. When $s \neq 0$, the tensors are constructed recursively as $(A^{(s)}[j]) = [A^{(s)}[j-1], C_j^s]$, where the matrix C_j^s has right dimension $\binom{j+d-2}{d-2}$ and all of its elements are either zero or one. The nonzero elements of the matrix $(C_j^s)_{\mu,\nu}$ are those whose components satisfy

$$\mu = \sum_{k=1}^{d-2} \sum_{n_k=0}^{(j-1)-\sum_{\ell=1}^k t_{\ell-1}} \binom{n_k + d - 2 - k}{d - 2 - k} + 1, \quad (\text{D7})$$

$$\begin{aligned} \nu &= \sum_{k=1}^{s-1} \sum_{n_k=0}^{j-\sum_{\ell=1}^k t_{\ell-1}} \binom{n_k + d - 2 - k}{d - 2 - k} \\ &+ \sum_{k=s}^{d-2} \sum_{n_k=0}^{j-\sum_{\ell=1}^k t_{\ell-2}} \binom{n_k + d - 2 - k}{d - 2 - k} + 1 \end{aligned} \quad (\text{D8})$$

for some choice of $\{t_n\}$ satisfying $\sum_n t_n = j - 1$. The tensors for the right half of the chain $j > \lceil \frac{N+1}{2} \rceil$ are constructed in the same way but with the role of column and row indices exchanged, i.e., they are the transpose.

This leaves a single site unaccounted for at $j = \lceil \frac{N+1}{2} \rceil$. All the coefficients that are needed to describe this operator can be placed in the tensor at this site. The $\mu\nu$ matrix element of $A^{(s)}[\lceil \frac{N+1}{2} \rceil]$ has the value $c_{\mathbf{t}}$ if

$$\mu = \binom{\|\mathbf{t}^{(L)}\|_1 - 2 + d}{d-1} + \sum_{k=1}^{d-2} \sum_{n_k=0}^{\sum_{\ell=k+1}^{d-1} t_{\ell}^{(L)} - 1} \binom{n_k + d - 2 - k}{d-2-k} + 1, \quad (\text{D9})$$

$$\nu = \binom{\|\mathbf{t}^{(R)}\|_1 - 2 + d}{d-1} + \sum_{k=1}^{d-2} \sum_{n_k=0}^{\sum_{\ell=k+1}^{d-1} t_{\ell}^{(R)} - 1} \binom{n_k + d - 2 - k}{d-2-k} + 1, \quad (\text{D10})$$

with $\mathbf{t}^{(L)} + \mathbf{t}^{(R)} + \hat{e}_s = \mathbf{t}$, where \hat{e}_s is the type vector with all elements equal to zero except for element s , which is equal to one and we use the convention that $\binom{n}{m} = 0$ if $m < 0$. The result of this construction is an exact MPS or MPO representation of the state or operator with bond dimension $\binom{\lceil \frac{N+1}{2} \rceil + d - 1}{d-1} = O(N^{d-1})$ for pure states and $\binom{\lceil \frac{N+1}{2} \rceil + d^2 - 1}{d^2 - 1} = O(N^{d^2-1})$ for operators.

APPENDIX E: PERMUTATION INVARIANT COMPILATION

Here we show that a polynomial depth circuit of permutation invariant unitaries can be compiled into a single permutation invariant unitary at a cost polynomial in system size. In Appendix D, we saw that a permutation invariant operator is specified by only $\binom{N+3}{N} = O(N^3)$ coefficients, one for each possible length-three-type vector for N sites. Thus, given two permutation invariant operators U and V with coefficients $a_{\mathbf{t}}$ and $b_{\mathbf{v}}$ respectively, our task is to find the coefficients $c_{\mathbf{v}'}$ of their product UV . We will prime the elements of the type vectors in the same way as the type vectors themselves so, for example, $\mathbf{t}'' = (t''_{(0,1)}, t''_{(1,0)}, t''_{(1,1)})$.

Ultimately, the resulting bra-ket pair on qubit j can end up with the form $|\mu\rangle\langle\nu|$ in only two ways. Either it emerged from a product of the form $|\mu\rangle\langle 0|0\rangle\langle\nu|$ or from a product of the form $|\mu\rangle\langle 1|1\rangle\langle\nu|$. It will be convenient to label the number of (μ, ν) pairs that emerge from the first type as $n_{\nu}^{\mu 0}$. Further, the contribution will be nonzero only if the resulting number of $t''_{(1,0)}$ pairs is equal to the number that emerges from the $|1\rangle\langle 0|0\rangle\langle 0|$ products plus the number that emerges from the $|1\rangle\langle 1|1\rangle\langle 0|$, that is, $t''_{(\mu,\nu)} = n_{\nu}^{\mu 0} + n_{\nu}^{\mu 1}$. We must also have $t_{(\mu,\nu)} = n_{\nu 0}^{\mu\nu} + n_{\nu 1}^{\mu\nu}$ and $t'_{(\mu,\nu)} = n_{\mu\nu}^{0\mu} + n_{\mu\nu}^{1\mu}$. This allows us to derive conditions on the type vector of the two input operators

in terms of these parameters and the type of output operator. For the first operator, we have

$$t_{(0,1)} = n_{10}^{01} + n_{11}^{01} = n_{10}^{01} + t'_{(0,1)} - n_{01}^{00}, \quad (\text{E1})$$

$$t_{(1,0)} = n_{00}^{10} + n_{01}^{10} = t'_{(1,0)} - n_{10}^{11} + n_{01}^{10}, \quad (\text{E2})$$

$$t_{(1,1)} = n_{10}^{11} + n_{11}^{11} = n_{10}^{11} + t'_{(1,1)} - n_{01}^{10}. \quad (\text{E3})$$

The conditions on the second operator type are

$$t'_{(0,1)} = n_{01}^{00} + n_{01}^{10}, \quad (\text{E4})$$

$$t'_{(1,0)} = n_{10}^{01} + n_{10}^{11}, \quad (\text{E5})$$

$$t'_{(1,1)} = n_{11}^{01} + n_{11}^{11} = t'_{(0,1)} - n_{01}^{00} + t'_{(1,1)} - n_{01}^{10}. \quad (\text{E6})$$

Thus, both input operators types are fixed in terms of the type of output operator and the four parameters $\{n_{01}^{00}, n_{01}^{10}, n_{10}^{01}, n_{10}^{11}\}$.

It is natural to first consider the product of the sum of all possible permutations of two of our representative elements of the equivalence classes, i.e.,

$$\left(\frac{1}{t_{(0,0)}! \prod_j t_j!} \right) \left(\frac{1}{t'_{(0,0)}! \prod_j t'_j!} \right) \times \sum_{\sigma, \sigma'} P_{\sigma} \hat{x}(\mathbf{t}) P_{\sigma}^{\dagger} P_{\sigma'} \hat{x}(\mathbf{t}') P_{\sigma'}^{\dagger}. \quad (\text{E7})$$

It is useful to note that the second line of this expression can be rewritten as

$$\sum_{\sigma \in S_N} P_{\sigma} \left(\hat{x}(\mathbf{t}) \sum_{\sigma' \in S_N} P_{\sigma^{-1}\sigma'} x(\mathbf{t}') P_{\sigma^{-1}\sigma'}^{\dagger} \right) P_{\sigma}^{\dagger} = \sum_{\sigma \in S_N} P_{\sigma} \left(\hat{x}(\mathbf{t}) \sum_{\tilde{\sigma} \in S_N} P_{\tilde{\sigma}} x(\mathbf{t}') P_{\tilde{\sigma}}^{\dagger} \right) P_{\sigma}^{\dagger}. \quad (\text{E8})$$

Each term in the sum in the parenthesis will be nonzero only if all of the zero-bras in $\hat{x}(\mathbf{t})$ line up with a zero-ket in $P_{\tilde{\sigma}} x(\mathbf{t}') P_{\tilde{\sigma}}^{\dagger}$ and similarly for the one-bras and one-kets. For a fixed values of the parameters $n_{\nu\rho}^{\mu\nu}$. The number of nonzero contributions are

$$\binom{t_{(0,0)}}{n_{01}^{00}} \binom{t_{(0,1)}}{n_{10}^{01}} \binom{t_{(1,0)}}{n_{01}^{10}} \binom{t_{(1,1)}}{n_{10}^{11}} = \frac{t_{(0,0)}! t_{(0,1)}! t_{(1,0)}! t_{(1,1)}!}{(n_{00}^{00}! n_{10}^{01}!) (n_{01}^{00}! n_{11}^{01}!) (n_{00}^{10}! n_{10}^{11}!) (n_{01}^{10}! n_{11}^{11}!)}. \quad (\text{E9})$$

The numerator exactly cancels the normalization associated with the type vector \mathbf{t} . To obtain the coefficients $c_{\mathbf{v}'}$, we will need to multiply this by the normalization associated with the type vector \mathbf{t}'' . The resulting factor is

$$\frac{t''_{(0,0)}! t''_{(0,1)}! t''_{(1,0)}! t''_{(1,1)}!}{(n_{01}^{00}! n_{10}^{01}!) (n_{01}^{10}! n_{10}^{11}!)} \binom{t_{(0,0)}}{n_{01}^{00}} \binom{t_{(0,1)}}{n_{10}^{01}} \binom{t_{(1,0)}}{n_{01}^{10}} \binom{t_{(1,1)}}{n_{10}^{11}} = \binom{t''_{(0,0)}}{n_{01}^{01}} \binom{t''_{(1,1)}}{n_{10}^{10}} \binom{t''_{(0,1)}}{n_{01}^{00}} \binom{t''_{(1,0)}}{n_{10}^{11}}. \quad (\text{E10})$$

The coefficients associated with UV are then

$$c_{\mathbf{v}'} = \sum_{n_{10}^{01}=0}^{t''_{(0,0)}} \sum_{n_{01}^{10}=0}^{t''_{(1,1)}} \sum_{n_{01}^{00}=0}^{t''_{(0,1)}} \sum_{n_{10}^{11}=0}^{t''_{(1,0)}} \binom{t''_{(0,0)}}{n_{10}^{01}} \binom{t''_{(1,1)}}{n_{01}^{10}} \binom{t''_{(0,1)}}{n_{01}^{00}} \binom{t''_{(1,0)}}{n_{10}^{11}} a_{(n_{10}^{01}+t''_{(0,1)}-n_{01}^{00}, t''_{(1,0)}-n_{10}^{11}+n_{01}^{10}, n_{10}^{11}+t''_{(1,1)}-n_{01}^{10})} b_{(n_{01}^{00}+n_{01}^{10}, n_{01}^{01}+n_{10}^{11}, t''_{(0,1)}-n_{01}^{00}+t''_{(1,1)}-n_{01}^{10})}. \quad (\text{E11})$$

These sums each run over $N + 1$ entries, so the complexity of computing this coefficient is $O(N^4)$. Since the output operator is specified by $O(N^3)$ coefficients, the complexity of the multiplication is $O(N^7)$ in time and $O(N^3)$ in memory. We emphasize that these coefficients contain all the information about how the operator acts on the entire state space, not just the symmetric subspace.

-
- [1] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [2] J. Kitching, S. Knappe, and E. A. Donley, Atomic sensors—a review, *IEEE Sens. J.* **11**, 1749 (2011).
- [3] D. Budker and M. Romalis, Optical magnetometry, *Nat. Phys.* **3**, 227 (2007).
- [4] S. A. Diddams, T. Udem, J. C. Bergquist, E. A. Curtis, R. E. Drullinger, L. Hollberg, W. M. Itano, W. D. Lee, C. W. Oates, K. R. Vogel, and D. J. Wineland, An optical clock based on a single trapped $^{199}\text{Hg}^+$ ion, *Science* **293**, 825 (2001).
- [5] A. Clairon, P. Laurent, G. Santarelli, S. Ghezali, S. Lea, and M. Bahoura, A cesium fountain frequency standard: Preliminary results, *IEEE Trans. Instrum. Meas.* **44**, 128 (1995).
- [6] J. H. Shirley, W. D. Lee, and R. E. Drullinger, Accuracy evaluation of the primary frequency standard NIST-7, *Metrologia* **38**, 427 (2001).
- [7] E. A. Donley, Nuclear magnetic resonance gyroscopes, in *SENSORS, 2010 IEEE, Waikoloa, HI* (IEEE, Piscataway, NJ, 2010), pp. 17–22.
- [8] A. K. Mohapatra, M. G. Bason, B. Butscher, K. J. Weatherill, and C. S. Adams, A giant electro-optic effect using polarizable dark states, *Nat. Phys.* **4**, 890 (2008).
- [9] L. Pezzè and A. Smerzi, Quantum phase estimation algorithm with Gaussian spin states, *PRX Quantum* **2**, 040301 (2021).
- [10] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Optimal frequency measurements with maximally correlated states, *Phys. Rev. A* **54**, R4649 (1996).
- [11] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, *Science* **306**, 1330 (2004).
- [12] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
- [13] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).
- [14] G. Tóth and I. Apellaniz, Quantum metrology from a quantum information science perspective, *J. Phys. A: Math. Theor.* **47**, 424006 (2014).
- [15] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, *Rev. Mod. Phys.* **90**, 035005 (2018).
- [16] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Spin squeezing and reduced quantum noise in spectroscopy, *Phys. Rev. A* **46**, R6797 (1992).
- [17] M. Kitagawa and M. Ueda, Squeezed spin states, *Phys. Rev. A* **47**, 5138 (1993).
- [18] N. Henkel, R. Nath, and T. Pohl, Three-dimensional roton excitations and supersolid formation in Rydberg-excited Bose-Einstein condensates, *Phys. Rev. Lett.* **104**, 195302 (2010).
- [19] B. Braverman, A. Kawasaki, E. Pedrozo-Peñañiel, S. Colombo, C. Shu, Z. Li, E. Mendez, M. Yamoah, L. Salvi, D. Akamatsu, Y. Xiao, and V. Vuletić, Near-unitary spin squeezing in ^{171}Yb , *Phys. Rev. Lett.* **122**, 223203 (2019).
- [20] I. Pogorelov, T. Feldker, C. D. Marciniak, L. Postler, G. Jacob, O. Kriegelsteiner, V. Podlesnic, M. Meth, V. Negnevitsky, M. Stadler, B. Höfer, C. Wächter, K. Lakhmanskiy, R. Blatt, P. Schindler, and T. Monz, Compact ion-trap quantum computing demonstrator, *PRX Quantum* **2**, 020343 (2021).
- [21] G. Pupillo, A. Micheli, M. Boninsegni, I. Lesanovsky, and P. Zoller, Strongly correlated gases of Rydberg-dressed atoms: Quantum and classical dynamics, *Phys. Rev. Lett.* **104**, 223002 (2010).
- [22] L. I. R. Gil, R. Mukherjee, E. M. Bridge, M. P. A. Jones, and T. Pohl, Spin squeezing in a Rydberg lattice clock, *Phys. Rev. Lett.* **112**, 103601 (2014).
- [23] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, Toward Heisenberg-limited spectroscopy with multiparticle entangled states, *Science* **304**, 1476 (2004).
- [24] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland, Creation of a six-atom ‘Schrödinger cat’ state, *Nature (London)* **438**, 639 (2005).
- [25] E. Davis, G. Bentsen, and M. Schleier-Smith, Approaching the Heisenberg limit without single-particle detection, *Phys. Rev. Lett.* **116**, 053601 (2016).
- [26] F. Fröwis, P. Sekatski, and W. Dür, Detecting large quantum Fisher information with finite measurement precision, *Phys. Rev. Lett.* **116**, 090801 (2016).
- [27] T. Macrì, A. Smerzi, and L. Pezzè, Loschmidt echo for quantum metrology, *Phys. Rev. A* **94**, 010102(R) (2016).
- [28] S. P. Nolan, S. S. Szigeti, and S. A. Haine, Optimal and robust quantum metrology using interaction-based readouts, *Phys. Rev. Lett.* **119**, 193601 (2017).
- [29] S. Colombo, E. Pedrozo-Peñañiel, A. F. Adiyatullin, Z. Li, E. Mendez, C. Shu, and V. Vuletić, Time-reversal-based quantum metrology with many-body entangled states, *Nat. Phys.* **18**, 925 (2022).
- [30] Z. Li, S. Colombo, C. Shu, G. Velez, S. Pilatowsky-Cameo, R. Schmied, S. Choi, M. Lukin, E. Pedrozo-Peñañiel, and

- V. Vuletić, Improving metrology with quantum scrambling *Science* **380**, 1381 (2023).
- [31] V. Bužek, R. Derka, and S. Massar, Optimal quantum clocks, *Phys. Rev. Lett.* **82**, 2207 (1999).
- [32] D. W. Berry and H. M. Wiseman, Optimal states and almost optimal adaptive measurements for quantum interferometry, *Phys. Rev. Lett.* **85**, 5098 (2000).
- [33] K. Macieszczak, M. Fraas, and R. Demkowicz-Dobrzański, Bayesian quantum frequency estimation in presence of collective dephasing, *New J. Phys.* **16**, 113002 (2014).
- [34] R. Kaubruegger, D. V. Vasilyev, M. Schulte, K. Hammerer, and P. Zoller, Quantum variational optimization of Ramsey interferometry and atomic clocks, *Phys. Rev. X* **11**, 041045 (2021).
- [35] N. Friis, D. Orsucci, M. Skotiniotis, P. Sekatski, V. Dunjko, H. J. Briegel, and W. Dür, Flexible resources for quantum metrology, *New J. Phys.* **19**, 063044 (2017).
- [36] S. Wölk, P. Sekatski, and W. Dür, Noisy distributed sensing in the Bayesian regime, *Quantum Sci. Technol.* **5**, 045003 (2020).
- [37] M. Jarzyna and R. Demkowicz-Dobrzański, True precision limits in quantum metrology, *New J. Phys.* **17**, 013010 (2015).
- [38] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, π -corrected Heisenberg limit, *Phys. Rev. Lett.* **124**, 030501 (2020).
- [39] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, 1980).
- [40] S. Personick, Application of quantum estimation theory to analog communication over quantum channels, *IEEE Trans. Inf. Theory* **17**, 240 (1971).
- [41] E. Martínez-Vargas, C. Pineda, F. Leyvraz, and P. Barberis-Blostein, Quantum estimation of unknown parameters, *Phys. Rev. A* **95**, 012136 (2017).
- [42] M. Tsang, Physics-inspired forms of the Bayesian Cramér-Rao bound, *Phys. Rev. A* **102**, 062217 (2020).
- [43] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, Optimized quantum sensing with a single electron spin using real-time adaptive measurements, *Nature Nanotech* **11**, 247 (2016).
- [44] T. Rosenband and D. R. Leibrandt, Exponential scaling of clock stability with atom number, [arXiv:1303.6357](https://arxiv.org/abs/1303.6357).
- [45] I. D. Leroux, N. Scharnhorst, S. Hannig, J. Kramer, L. Pelzer, M. Stepanova, and P. O. Schmidt, On-line estimation of local oscillator noise and optimisation of servo parameters in atomic clocks, *Metrologia* **54**, 307 (2017).
- [46] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Improvement of frequency standards with quantum entanglement, *Phys. Rev. Lett.* **79**, 3865 (1997).
- [47] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, *Nat. Phys.* **7**, 406 (2011).
- [48] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, The elusive Heisenberg limit in quantum-enhanced metrology, *Nat. Commun.* **3**, 1063 (2012).
- [49] M. Jarzyna and R. Demkowicz-Dobrzański, Matrix product states for quantum metrology, *Phys. Rev. Lett.* **110**, 240405 (2013).
- [50] J. Kołodyński and R. Demkowicz-Dobrzański, Efficient tools for quantum metrology with uncorrelated noise, *New J. Phys.* **15**, 073043 (2013).
- [51] M. Tsang, Quantum metrology with open dynamical systems, *New J. Phys.* **15**, 073005 (2013).
- [52] K. Yamamoto, S. Endo, H. Hakoshima, Y. Matsuzaki, and Y. Tokunaga, Error-mitigated quantum metrology via virtual purification, *Phys. Rev. Lett.* **129**, 250503 (2022).
- [53] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [54] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [55] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem, [arXiv:1412.6062](https://arxiv.org/abs/1412.6062).
- [56] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [57] R. Kaubruegger, P. Silvi, C. Kokail, R. van Bijnen, A. M. Rey, J. Ye, A. M. Kaufman, and P. Zoller, Variational spin-squeezing algorithms on programmable quantum sensors, *Phys. Rev. Lett.* **123**, 260505 (2019).
- [58] Q. Zhuang and Z. Zhang, Physical-layer supervised learning assisted by an entangled sensor network, *Phys. Rev. X* **9**, 041023 (2019).
- [59] B. Koczor, S. Endo, T. Jones, Y. Matsuzaki, and S. C. Benjamin, Variational-state quantum metrology, *New J. Phys.* **22**, 083038 (2020).
- [60] L. Pezzè and A. Smerzi, Heisenberg-limited noisy atomic clock using a hybrid coherent and squeezed state protocol, *Phys. Rev. Lett.* **125**, 210503 (2020).
- [61] M. Schulte, V. J. Martínez-Lahuerta, M. S. Scharnagl, and K. Hammerer, Ramsey interferometry with generalized one-axis twisting echoes, *Quantum* **4**, 268 (2020).
- [62] X. Yang, J. Thompson, Z. Wu, M. Gu, X. Peng, and J. Du, Probe optimization for quantum metrology via closed-loop learning control, *npj Quantum Inf.* **6**, 62 (2020).
- [63] X. Yang, X. Chen, J. Li, X. Peng, and R. Laflamme, Hybrid quantum-classical approach to enhanced quantum metrology, *Sci. Rep.* **11**, 672 (2021).
- [64] J. J. Meyer, J. Borregaard, and J. Eisert, A variational toolbox for quantum multi-parameter estimation, *npj Quantum Inf.* **7**, 89 (2021).
- [65] Z. Ma, P. Gokhale, T. Zheng, S. Zhou, X. Yu, L. Jiang, P. Maurer, and F. T. Chong, Adaptive circuit learning for quantum metrology, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Broomfield, CO (IEEE, Piscataway, NJ, 2021).
- [66] A. Altherr and Y. Yang, Quantum metrology for non-Markovian processes, *Phys. Rev. Lett.* **127**, 060501 (2021).
- [67] Y. Xia, W. Li, Q. Zhuang, and Z. Zhang, Quantum-enhanced data classification with a variational entangled sensor network, *Phys. Rev. X* **11**, 021047 (2021).
- [68] S. C. Carrasco, M. H. Goerz, Z. Li, S. Colombo, V. Vuletic, V. S. Malinovsky, Extreme spin squeezing via optimized one-axis twisting and rotations, *Phys. Rev. Appl.* **17**, 064050 (2022).
- [69] C. D. Marciniak, T. Feldker, I. Pogorelov, R. Kaubruegger, D. V. Vasilyev, R. van Bijnen, P. Schindler, P. Zoller, R. Blatt, and T. Monz, Optimal metrology with programmable quantum sensors, *Nature (London)* **603**, 604 (2022).

- [70] J. L. Beckey, M. Cerezo, A. Sone, and P. J. Coles, Variational quantum algorithm for estimating the quantum Fisher information, *Phys. Rev. Res.* **4**, 013083 (2022).
- [71] T. Zheng, A. Li, J. Rosen, S. Zhou, M. Koppenhfer, Z. Ma, F. T. Chong, A. A. Clerk, L. Jiang, and P. C. Maurer, Preparation of metrological states in dipolar-interacting spin systems, *npj Quantum Inf.* **8**, 150 (2022).
- [72] L. B. Ho, A stochastic evaluation of quantum Fisher information matrix with generic Hamiltonians *EPJ Quantum Technol.* **10**, 37 (2023).
- [73] R. Liu, Z. Wu, X. Yang, Y. Li, H. Zhou, Y. Chen, H. Yuan, X. Peng, and J. Du, Variational quantum metrology with Loschmidt echo, [arXiv:2211.12296](https://arxiv.org/abs/2211.12296).
- [74] C. M. Caves, Quantum-mechanical noise in an interferometer, *Phys. Rev. D* **23**, 1693 (1981).
- [75] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [76] K. Macieszczak, Quantum fisher information: Variational principle and simple iterative algorithm for its efficient computation, [arXiv:1312.1356](https://arxiv.org/abs/1312.1356).
- [77] K. Chabuda, J. Dziarmaga, T. J. Osborne, and R. Demkowicz-Dobrzański, Tensor-network approach for quantum metrology in many-body quantum systems, *Nat. Commun.* **11**, 250 (2020).
- [78] H. L. van Trees, K. L. Bell, and Z. Tian, *Detection, Estimation, and Modulation Theory* (Wiley, Hoboken, NJ, 2013).
- [79] R. D. Gill and B. Y. Levit, Applications of the van Trees inequality: A Bayesian Cramér-Rao bound, *Bernoulli* **1**, 59 (1995).
- [80] M. A. Paris, Quantum estimation for quantum technology, *Int. J. Quantum. Inform.* **07**, 125 (2009).
- [81] J. Lee, G. Vrijsen, I. Teper, O. Hosten, and M. A. Kasevich, Many-atom cavity qed system with homogeneous atom cavity coupling, *Opt. Lett.* **39**, 4005 (2014).
- [82] A. Kawasaki, B. Braverman, E. Pedrozo-Peñafiel, C. Shu, S. Colombo, Z. Li, Ö. Özel, W. Chen, L. Salvi, A. Heinz, D. Levonian, D. Akamatsu, Y. Xiao, and V. Vuletić, Geometrically asymmetric optical cavity for strong atom-photon coupling, *Phys. Rev. A* **99**, 013437 (2019).
- [83] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, *Quantum Inform. Comput.* **7**, 401 (2007).
- [84] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Ann. Phys.* **349**, 117 (2014).
- [85] J. C. Bridgeman and C. T. Chubb, Hand-waving and interpretive dance: An introductory course on tensor networks, *J. Phys. A: Math. Theor.* **50**, 223001 (2017).
- [86] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, *Rev. Mod. Phys.* **93**, 045003 (2021).
- [87] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.* **7**, 308 (1965).
- [88] M. J. Box, A new method of constrained optimization and a comparison with other methods, *Comput. J.* **8**, 42 (1965).
- [89] J. A. Richardson and J. L. Kuester, Algorithm 454: The complex method for constrained optimization [e4], *Commun. ACM* **16**, 487 (1973).
- [90] D. Kraft, A software package for sequential quadratic programming, Institut für Dynamik der Flugsysteme technical report, 1988.
- [91] D. Kraft, Algorithm 733: Tomp–fortran modules for optimal control calculations, *ACM Trans. Math. Software* **20**, 262 (1994).
- [92] S. G. Johnson, The NLOpt nonlinear-optimization package, <http://github.com/stevengj/nlopt>.
- [93] M. Innes, Don't unroll adjoint: Differentiating SSA-form programs, [arXiv:1810.07951](https://arxiv.org/abs/1810.07951).
- [94] Fastgaussquadrature.jl, <https://github.com/JuliaApproximation/FastGaussQuadrature.jl>.
- [95] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, *SciPost Phys. Codebases* **4** (2022).
- [96] M. Fishman, S. R. White, and E. M. Stoudenmire, Codebase release 0.3 for ITensor, *SciPost Phys. Codebases* **4** (2022).