Optimal universal quantum error correction via bounded reference frames

Yuxiang Yang [1,2,*] Yin Mo,[2] Joseph M. Renes [1] Giulio Chiribella,[2,3,4,5] and Mischa P. Woods[1]

[1]*Institute for Theoretical Physics, ETH Zürich, Switzerland*
[2]*QICI Quantum Information and Computation Initiative, Department of Computer Science,*
The University of Hong Kong, Pokfulam Road, Hong Kong
[3]*Department of Computer Science, Parks Road, Oxford OX1 3QD, United Kingdom*
[4]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[5]*The University of Hong Kong Shenzhen Institute of Research and Innovation, 5/F, Key Laboratory Platform Building,*
No.6, Yuexing 2nd Rd., Nanshan, Shenzhen 518057, China

Error correcting codes with a universal set of transversal gates are a desideratum for quantum computing. Such codes, however, are ruled out by the Eastin-Knill theorem. Moreover, the theorem also rules out codes which are covariant with respect to the action of transversal unitary operations forming continuous symmetries. In this work, starting from an arbitrary code, we construct *approximate* codes which are covariant with respect to the entire group of local unitary gates in dimension d ($<\infty$), using quantum reference frames. We show that our codes are capable of efficiently correcting different types of erasure errors. When only a small fraction of the n qudits upon which the code is built are erased, our covariant code has an error that scales as $1/n^2$, which is reminiscent of the Heisenberg limit of quantum metrology. When every qudit has a chance of being erased, our covariant code has an error that scales as $1/n$. We show that the error scaling is optimal in both cases. Our approach has implications for fault-tolerant quantum computing, reference frame error correction, and the AdS-CFT duality.

I. INTRODUCTION

Reliable universal quantum computation requires fault-tolerant error correction [1], the ability to correct errors with gates that are themselves noisy. Achieving such quantum error correcting codes (QECCs) is a notoriously challenging task, due to fundamental limitations such as quantum no cloning.

One of the earliest proposals to achieve fault-tolerance for universal quantum computation was the idea of implementing all the logical gates "transversally" [2,3], which is the following idea: Given a qudit logical space and an n-qudit physical space, find an encoder $\mathcal{E}_{\mathrm{cov}}$ which maps all logical gates $V \in \mathsf{SU}(d)$, the group of unitaries in d dimensions, to a tensor product of physical gates:

$$\mathcal{E}_{\mathrm{cov}} \circ \mathcal{V}_{\mathrm{L}} = (\mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_n) \circ \mathcal{E}_{\mathrm{cov}}, \qquad (1)$$

where $\mathcal{V}_{\mathrm{L}}(\cdot) = V(\cdot)V^{\dagger}$ is the unitary channel corresponding to logical gate V_{L}, and $\mathcal{V}_m$ ($m = 1, \ldots, n$) is either the unitary channel $V(\cdot)V^{\dagger}$ or the identity channel, on the mth physi-

cal qudit.[1] Therefore, in such proposals, all gates required to achieve universal computation would be realized at the physical level by applying a tensor product of local gates. This structure is naturally suited to error correction, since an error in one of the physical qudits does not easily propagate to other physical qudits—keeping errors local so that it can be efficiently corrected afterwards.

However, as shown by Easting and Knill, there cannot exist a code $\mathcal{E}$ satisfying Eq. (1), with a finite dimensional code space that *perfectly* corrects single-qudit errors [4]. Moreover, their result actually holds more generally for any set of qudit gates which form a continuous symmetry, that is to say, any set of gates $\mathcal{V}$ which act on qudits and form a Lie subgroup of $\mathsf{SU}(d)$ [4–6].

The code space is simply the image, in the physical space, of the encoding map, while a code that perfectly corrects is simply one for which the decoder $\mathcal{D}_{\mathrm{cov}}$ can correct any error $\mathcal{C}$ and still decode perfectly any logical state ρ_{L}:

$$\mathcal{D}_{\mathrm{cov}} \circ \mathcal{C} \circ \mathcal{E}_{\mathrm{cov}} \circ \rho_{\mathrm{L}} = \rho_{\mathrm{L}}. \qquad (2)$$

The Eastin-Knill theorem does not rule out the ability to correct local errors for a nonuniversal set of gates, however.

---

*mischa@phys.ethz.ch

---

[1]One may feel that Eq. (1) should be relaxed to hold only for a finite, universal gate set. However, observe that since the transversality condition is preserved under gate composition, and the set of gates generated by a universal gate set is dense in $\mathsf{SU}(d)$, these two conditions are effectively equivalent.

A case in point is the set of Clifford gates, which can be implemented transversally but lack one crucial gate in order to form a universal set. Supplementing this set with the missing gate is the idea behind one of the frontrunner proposals for universal quantum computation, using so-called "magic states" [7]. Several other schemes to relax the transversality condition also exist [8–14].

In this work, we demonstrate a different approach to circumvent the Eastin-Knill theorem. In particular, all gates can be applied transversally under various local error models in our scheme. The key ingredients are quantum reference frames and randomness. We now introduce the former before explaining their relevance to our scheme.

In physics, all observations are reported relative to a reference frame. While the reference frames have traditionally been *treated according to the laws of classical physics*, the usage of quantum states to encode reference frame information (a Cartesian coordinate system, for instance) in quantum superpositions has been shown to be advantageous in problems involving reference frame alignment and overcoming superselection rules [15–21]. Quantum reference frames also play a crucial role in demystifying a number of controversies and paradoxes [22–25], in addition to unifying various different paradigms [26]. There has been a renaissance very recently, due to generalization of quantum reference frame transformations to a "superposition of coordinate transformations," which consistently describe the physics without appealing to an external, absolute reference frame [27–33].

Quantum reference frames in the context of QECCs was first explored in Ref. [34], where classical idealized reference frames[2] were employed. The hypothetical setup allows for exact decoding, and does not violate the Eastin-Knill theorem due to the use of an infinite dimensional code space. While impractical due to infinite dimensionality, the approach in Ref. [34] paved the way for another approach to circumvent the Eastin-Knill theorem: using finite dimensional reference frames and decoders which only recover approximately. This route was followed in Ref. [5], where finite quantum reference frames have been used to allow for a single Abelian family of transversal gates. This construction, while useful, does not allow for universal quantum computation with transversal gates.

Here we further develop this approach and design a new family of quantum reference frames to achieve QECC constructions with all logical gates being transversal. Quantum reference frames have an inherent quantum uncertainty which unavoidably leads to a small error in the decoding. Consequently, our quantum error correcting code is only approximate. On a theoretical level, this is essential to circumvent the Eastin-Knill theorem while achieving our objective of a universal quantum gate set. On a practical level, this error can be made smaller than any chosen tolerance by increasing the size of the quantum reference frame.

---

[2]An idealized reference frame, also know as a "perfect" reference frame, is one whose orientation can be deterministically obtained via measurement which is contrary to a quantum reference frame whose orientation is subject to quantum uncertainty. See Ref. [18] for definition and Ref. [5] for further insight.
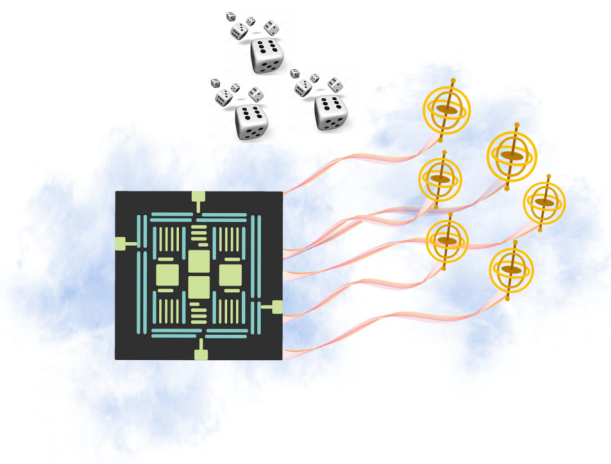

FIG. 1. Error correction via reference frames. The physical space of the computer is over n qudits which are divided into two subspaces: the computational space (black and green) and a reference frame space (yellow and brown). The former is where the gates needed for the algorithm are applied transversally, while the latter is where the reference frames are located. The two systems are only classically correlated at any given moment in our protocols.

The physical space necessary for our implementation of the QECC consists of two parts, the *computational space* and the *reference frame space*. Roughly speaking, the computational space is where the logical information is stored and the sequence of logical gates needed for the computation are applied on various copies. The reference frames on the other hand, play the role of recording information about which gate was applied—analogously to how gyroscopes record a Cartesian coordinate system. See Fig. 1.

At an abstract level, our encoder works by choosing a quantum reference frame which can record the "coordinates" of any logical state. The quantum reference frame states are approximately distinguishable by measurements under transformations of $\mathsf{SU}(d)$—the special group of unitary transformations in d dimensions. To implement any logical gate under this encoding, we can simply apply the gate transversally on the computational space in tandem with updating our reference frames with a corresponding transformation.

However, before this paradigm can be implemented, one needs to perform the encoding—this is where the randomness comes into play. We pick a logical gate at random, apply it transversally to the computational system and a copy to the reference frame. The reference frame information is retrieved during the decoding stage via measurement.

In the regime of error correction via finite quantum reference frames, the error of approximate recovery is dependent on the size of the reference frames, leading to a trade-off relation. It is thus of prime importance to identify the optimal scaling of the approximation error with respect to the size of the reference frame, so that the cost of making the recovery errors admissible can be determined. Here we analyze the error scaling of our QECC construction in two different commonly studied error models and show that our approach is optimal in both cases. Besides the analytical analysis, we also conduct numerical simulation that further justifies our analysis in various practical error models.

In a nutshell, our contribution is to show that the use of quantum reference frames allows for approximate QECC where all gates necessary for universal quantum computation can be applied transversally, and to identify the optimal tradeoff between the approximation error and the size of the reference frames. This was the initial idea of the pioneers of quantum error correction [2,3], who, in light of the Eastin-Knill theorem, invented distinct methods alternative to finding a transversal gate set for quantum error correction. Our work thus offers a new test bed for quantum error correction whose full potential can now be further explored. An immediate question is whether the encoder and the decoder in our construction can be implemented fault-tolerantly. As discussed in more details later, there are several paths that potentially lead to this ultimate purpose, though the concrete approach is still unknown. This is thus an important open question left for investigation in future research.

II. RESULTS

Overview. We start by introducing the relevant definitions to define our encoder, decoder, and characterization of the decoding error. We then consider two errors models: the *i.i.d. error model* and the *weak error model* and fully characterize the performance of our QECC under these different scenarios. Details appear in the appendices.

Construction of covariant encoder and decoder. A QECC is characterized by an encoder-decoder pair $(\mathcal{E}, \mathcal{D})$. The encoder $\mathcal{E}$ is a quantum channel mapping a logical qudit, whose dimension we denote by d, to a state of a physical register consisting of several qudits. The decoder $\mathcal{D}$ corrects possible errors in the physical register and maps the state back to a logical qudit. Since our protocol uses reference frames and randomness to convert an arbitrary encoder $\mathcal{E}$ and decoder $\mathcal{D}$ to a covariant encoder-decoder pair, we refer to $\mathcal{E}$ and $\mathcal{D}$ as the *subroutine* encoder and decoder. The n-qudit physical register is divided into two parts, the *computational register* C, consisting in n_C qudits and a *reference frame register* R, consisting in the remaining $n_R = n - n_C$ qudits. The subroutine decoder, $\mathcal{D}$, maps to the computational register only.

In order to be resilient to errors, the reference frame register may consist of $s_R \geqslant 1$ identical blocks, each in the quantum state $|\psi\rangle$. Here we choose $|\psi\rangle$ to be an entangled state of a bipartite system $\mathcal{H} \otimes \mathcal{H}'$ ($\mathcal{H}' \simeq \mathcal{H}$), as they offer superior performance [17,35,36]. Each part of the bipartite system further consists of several qudits. This reference frame is a finite dimensional quantum state $|\psi\rangle$. Rotating every qudit in the part of $|\psi\rangle$ on $\mathcal{H}$ by the same (unknown) unitary U in $\mathsf{SU}(d)$ (while keeping the part on $\mathcal{H}'$ unchanged) yields a state that, when measured appropriately, allows one to deduce U with high accuracy. The better the quality of $|\psi\rangle$, the smaller the uncertainty in U is.

Our covariant encoder is defined as

$$\mathcal{E}_{\text{cov}}(\cdot) = \int dU\, \mathcal{U}_C \circ \mathcal{E} \circ \mathcal{U}_L^{-1}(\cdot) \otimes \mathcal{U}_R(\Psi), \qquad (3)$$

where $\Psi := |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = |\psi\rangle^{\otimes s_R}$ is the initial state of the reference frame register and dU is the Haar measure on $\mathsf{SU}(d)$. The unitary channels in Eq. (3) are defined by $\mathcal{U}_C := \mathcal{U}^{\otimes n_C}$ and $\mathcal{U}_L := \mathcal{U}$ the actions of an element $U \in \mathsf{SU}(d)$ on

the computational register and the logical register, and by $\mathcal{U}_R := (\mathcal{U} \otimes \mathcal{I})^{\otimes n_R/2}$ its action on the reference frame register (as one part of the reference frame state is rotated and the other part remains unchanged), where $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$. See Fig. 2 for an example of how to construct the encoder.

The covariant decoder consists of first measuring the reference frame with a POVM $\{M_{\hat{U}}\, d\hat{U}\}$, which yields an estimate $\hat{U}$, and then applying $\hat{\mathcal{U}}_C^{-1}$. Here $d\hat{U}$ is again the Haar measure. We then perform the subroutine decoder $\mathcal{D}$ of the original code to correct errors, and finally we redo $\hat{\mathcal{U}}_L$ to obtain the final, decoded output. This gives rise to the channel

$$\mathcal{D}_{\text{cov}}(\cdot) = \int d\hat{U}\, (\hat{\mathcal{U}}_L \circ \mathcal{D} \circ \hat{\mathcal{U}}_C^{-1} \otimes \mathcal{M}_{\hat{U}})(\cdot), \qquad (4)$$

where the map $\mathcal{M}_{\hat{U}}(\cdot) := \text{Tr}[(\cdot)M_{\hat{U}}]$ outputs the probability density function of obtaining measurement outcome estimate $\hat{U}$.

One may be concerned about the practicality of constructing Eqs. (3) and (4), since sampling from the Haar measure is essentially to randomly apply all unitary gates, which might be computationally hard. In fact, we show later that, using the technique of unitary t-designs [37–44], we only require to sample from a discrete set of unitary gates in a computationally efficient manner.

Errors arise in this procedure both from any noise that occurs between encoder-decoder, as well as in the measurement of the reference frame itself. Nevertheless, we show that the overall error vanishes with the number of qudits upon which we construct the covariant code.

As promised, the encoder in Eq. (3) satisfies the transversality condition, Eq. (1), thanks to the well-known invariance properties of the Haar measure. In our particular case, it takes the form:

$$\mathcal{E}_{\text{cov}} \circ \mathcal{V}_L = (\mathcal{V}_C \otimes \mathcal{V}_R) \circ \mathcal{E}_{\text{cov}} \qquad (5)$$

for any $V \in \mathsf{SU}(d)$, where $\mathcal{V}_C := \mathcal{V}^{\otimes n_C}$ and $\mathcal{V}_L := \mathcal{V}$, $\mathcal{V}_R := (\mathcal{V} \otimes \mathcal{I})^{\otimes n_R/2}$ with $\mathcal{V} := V(\cdot)V^\dagger$, and $\mathcal{I}$ the identity channel. Notice that since our decoder is also covariant, i.e. $\mathcal{D}_{\text{cov}} \circ (\mathcal{V}_C \otimes \mathcal{V}_R) = \mathcal{V}_L \circ \mathcal{D}_{\text{cov}}$, our entire code is covariant: $\mathcal{D}_{\text{cov}} \circ \mathcal{E}_{\text{cov}} \circ \mathcal{V}_L = \mathcal{V}_L \circ \mathcal{D}_{\text{cov}} \circ \mathcal{E}_{\text{cov}}$.

Reference frame structure. There is a trade-off: the optimal pair of the initial reference frame state Ψ and the POVM $\{M_{\hat{U}}\, d\hat{U}\}$, which yields the smallest decoding errors when no noise is present, tends not to be very resilient to noise. Conversely, those which are highly resilient to noise, tend to lead to substantial decoding errors. As such, the optimal reference frame choice depends on the noise model. We will only provide the general structure considered here, and specialize later as specific noise models are introduced.

We divide the n_R reference frame register qudits into s_R blocks of identical systems, i.e., $\mathcal{H}_R = \mathcal{H}_{R,1} \otimes \mathcal{H}_{R,2} \otimes \cdots \otimes \mathcal{H}_{R,s_R}$ and construct identical states on each of them, $\Psi = \psi^{\otimes s_R}$. Each block consists in $2m$ qudits each, such that $2m\, s_R = n_R$. We assume that the error occurring on each of the reference frames is detectable, i.e., it acts as follows for all k: it takes the state on $\mathcal{H}_{R,k}$ to an error space $\mathcal{H}_{\text{err},k}$ that is orthogonal to $\mathcal{H}_{R,k}$. This is the case, for instance, for erasure errors.

The measurement performed on the reference frame during the decoding stage, is a two step process: We first perform

(a)

$\mathcal{U}^{-1} \circ \rho_{\rm L}$

1 qudit

$\mathcal{U}(\cdot) = U(\cdot)U^\dagger$

(b)

$\mathcal{E} \circ \mathcal{U}^{-1} \circ \rho_{\rm L}$

$\mathcal{U}_{\rm C} = \mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U}$

$n_{\rm C}$ qudits

(c)

$n_{\rm C}$ qudits $n_{\rm R}$ qudits

(d)

$(\mathcal{I} \otimes \mathcal{U})^{\otimes m}$
$(\mathcal{I} \otimes \mathcal{U})^{\otimes m}$ $(\mathcal{I} \otimes \mathcal{U})^{\otimes m}$

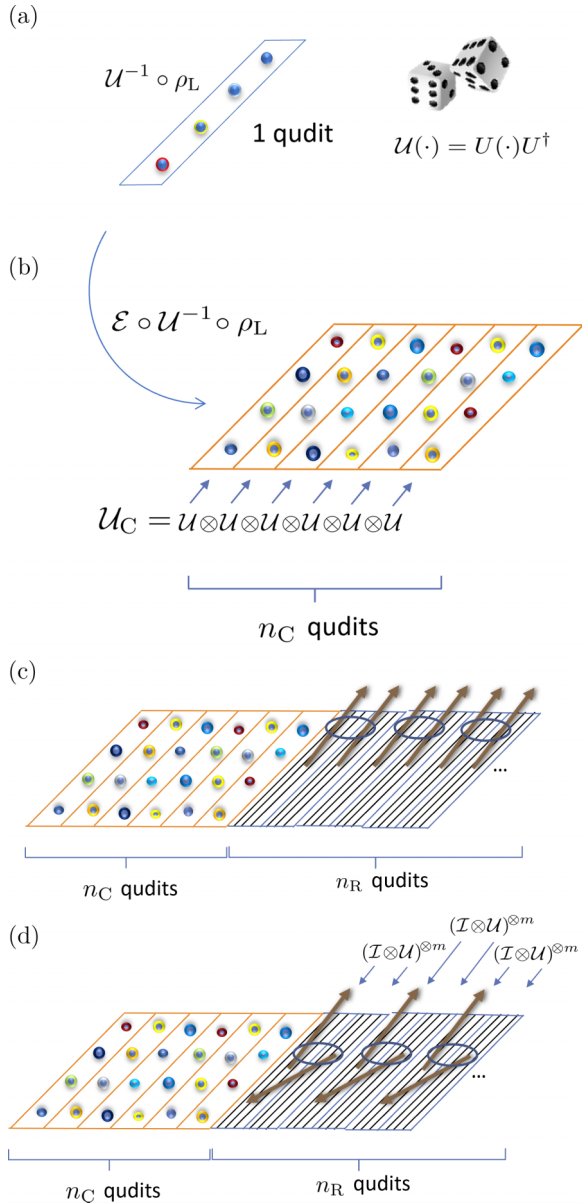$n_{\rm C}$ qudits $n_{\rm R}$ qudits

FIG. 2. Covariant encoder construction. (The following is an in-principle method to implement our QECC. For practical implementations, see section *Towards fault-tolerant QECCs*.) Given an arbitrary channel $\mathcal{E}$ encoding one logical qudit into $n_{\rm C}$ qudits, and a corresponding decoder $\mathcal{D}$, our covariant encoding protocol runs as follows. (a) Choose a unitary $U \in {\sf SU}(d)$ at random (in practice, this can be achieved by sampling from a large enough finite set of gates) and apply its inverse unitary channel $\mathcal{U}_{\rm L}^{-1}$ to the logical state $\rho_{\rm L}$ one wishes to encode. (b) Perform $\mathcal{U}_{\rm C} \circ \mathcal{E}$, where $\mathcal{U}_{\rm C}$ is the unitary channel corresponding to the representation of U in the computational space, and $\mathcal{U}_{\rm C}$ are local gates which will be transversal for the new code. (c) Such an operation alone does not result in a useful code, as the identification of the particular U is needed to later correct errors and decode. Thus, to store U efficiently, we append $n_{\rm R}$ quantum reference frames, and (d) apply it to a quantum reference frame, resulting in the unitary's identity being encoded into the reference frame's state. The outcome of steps 1 to 4 is a new code with a covariant encoder $\mathcal{E}_{\rm cov}$ which outputs an encoded state on the original computational and reference frame registers, C and R.

binary projective measurements onto $\mathcal{H}_{{\rm R},k}$ and $\mathcal{H}_{{\rm err},k}$ for all blocks $k = 1, \ldots, s_{\rm R}$, followed by discarding those blocks whose state is in the error space. We then measure the remaining blocks to obtain an estimate $\hat{U}$, which is used to decode the information in the computational space [see Eq. (4)].

Decoding error and noise structure. There are two contributing factors to errors in our covariant scheme: noise occurring between encoding and decoding, and errors in the decoding due to the finiteness of the reference frames. Here we show that our protocol is able to recover from both of these. Since these errors hinder our ability to decode, we refer to them as *decoding errors* and quantify it via the diamond norm:

$$\epsilon_{\rm cov} := \max_{V \in {\sf SU}(d)} \tfrac{1}{2} \| \mathcal{D}_{\rm cov} \circ \mathcal{C} \circ (\mathcal{V}_{\rm C} \otimes \mathcal{V}_{\rm R}) \circ \mathcal{E}_{\rm cov} - \mathcal{V}_{\rm L} \|_\diamond, \quad (6)$$

where $\mathcal{C}$ is the noise channel being considered, and the subscript in $\epsilon_{\rm cov}$ is to remind the reader that the error is originating from making an encoder-decoder pair covariant. We assume it to be a statistical mixture over channels on the computation and reference frame registers, i.e.,

$$\mathcal{C} = \sum_j p_j \mathcal{C}_{j,{\rm C}} \otimes \mathcal{C}_{j,{\rm R}}, \quad (7)$$

where $\mathcal{C}_{j,{\rm C/R}}$ are channels characterizing the error on the computational/reference frame register, and $\{p_j\}$ is a probability distribution. We also need to assume that the errors are covariant, namely $\mathcal{C} \circ (\mathcal{V}_{\rm C} \otimes \mathcal{V}_{\rm R}) = (\mathcal{V}_{\rm C} \otimes \mathcal{V}_{\rm R}) \circ \mathcal{C}$ for all $V \in {\sf SU}(d)$. For example, erasure errors are of this form.

In Appendix A, we show that the total decoding error satisfies a bound of the form $\epsilon_{\rm cov} \leqslant \sum_j p_j C_j$, where C_j only depends on either the computational error $\mathcal{C}_{j,{\rm C}}$ or the reference frame error $\mathcal{C}_{j,{\rm R}}$; whichever error is dominant.

Error bounds for the weak error model. Our first result concerns the case in which relatively few errors occur. Specifically, the error is that at most $n_{\rm e}$ qudits among the n qudit systems composing the computational register and the reference frame register, are randomly lost:

$$\mathcal{C} = \sum_{\sf s} p_{\sf s} (\mathcal{C}_{\rm e})_{\sf s}, \quad (8)$$

where the summation runs over all subsets ${\sf s} \subset \{1, 2, \ldots, n\}$ of cardinality at most $n_{\rm e}$, and $(\mathcal{C}_{\rm e})_{\sf s}$ denotes the erasure of qudits whose labels are in the set ${\sf s}$, and $\{p_{\sf s}\}$ is a probability distribution.

Any quantum error-correcting code over qudits of distance at least $k + 1$ can perfectly correct k erasures [3]. For instance, the polynomial codes of Aharonov and Ben-Or are $[2k + 1, 1, k + 1]_d$ stabilizer codes with this property [45]. In these cases, we employ one of the perfect codes as the noncovariant subroutine $(\mathcal{E}, \mathcal{D})$ of our code. Since we only need to encode one logical qudit, the perfect code requires only $n_{\rm C} = O(1)$ computational qudits.

This is to say: the subroutine code $(\mathcal{E}, \mathcal{D})$ we use requires only $O(1)$ computational qudits and corrects perfectly up to $n_{\rm e}$ erasure errors on the computational register. Next we arrange the reference frame to perform well against this type of noise. To this purpose, we set $s_{\rm R} = n_{\rm e} + 1$, so that the reference frames has $n_{\rm e} + 1$ blocks. Crucially, these conditions imply

that at least one of the reference frame states will survive the erasure, and we can measure it to obtain the embedded rotation.

On each block we construct a reference frame state. Since the noise is weak, we can utilize as much entanglement as possible to enhance the accuracy of the reference frame decoding. To this purpose, we design a new family of highly coupled quantum reference frame states, which turn out to be a highly nontrivial generalization of the sine-shape states [46], and thus we name them *generalized sine states*. Using the generalized sine states as reference frames for our error correction protocol, we can recover the information with high accuracy. We derive an upper bound on the recovery error ϵ_{cov} in our protocol for the weak error model, defined by Eq. (8):

$$\epsilon_{\mathrm{cov}} \leqslant \frac{81\pi^2 d^4 (d-1)^2 (n_{\mathrm{e}}+1)^2 (n_{\mathrm{C}}+d-1)^2}{2n_{\mathrm{R}}^2} + O(n_{\mathrm{R}}^{-3}).$$

(9)

All details can be found in Appendix B. Since n_{C} can be chosen to be $O(1)$ and $n_{\mathrm{R}} = n - n_{\mathrm{C}}$ is close to n, our protocol achieves the Heisenberg limit $1/n^2$ with respect to the total number of qudit systems.

It is most challenging to correct the errors in the weak error model when the locations of the erasures are most uncertain, that is, when $p_0 = 0$, $|\mathbf{s}| = n_{\mathrm{e}}$ and the $p_{\mathbf{s}}$ are constant, in Eq. (8). In this worse case scenario, we can lower bound the decoding error:

$$\epsilon_{\mathrm{cov}} \geqslant \frac{1}{16n^2(1+1/n_{\mathrm{e}})}.$$

(10)

What is more, we show that *any* encoder-decoder pair which satisfies the transversality condition Eq. (1), encoding one qudit into n physical qudits which is subject to the weak error model Eq. (8) with maximum uncertainty as described above, has a decoding error which is lower bounded by Eq. (10). As shown in Appendix B, the derivation of Eq. (10) comes from a strengthening of a bound found in Ref. [47] (see also Ref. [48] for a more recent result with similar bounds). Since the bound Eq. (10) matches the performance of our protocol in scaling, we conclude that the optimal error scaling of covariant codes is $1/n^2$.

Error bounds for the i.i.d. errorm Model. We now turn our attention to types of erasure errors which are stronger than the one considered in the previous section and more realistic. The error affects each qudit independently, erasing the jth qudit with a probability p_{e}:

$$\mathcal{C} = \bigotimes_{j=1}^{n} ((1-p_{\mathrm{e}})\mathcal{I} + p_{\mathrm{e}}\mathcal{C}_{\mathrm{e}}) \quad p_{\mathrm{e}} \in \left(0, \tfrac{1}{2}\right),$$

(11)

where $\mathcal{C}_{\mathrm{e}}$ denotes the single-qudit erasure channel. This error model can be written in the form Eq. (7), and is thus compatible with our prior assumptions. Since the erasure channel is degradable, if $p_{\mathrm{e}} \geqslant 1/2$ the information leaked to the environment would not be retrievable. Note that, in general, the error model on each qudit does not have to be identical, and each qudit j can have distinct probability $p_{\mathrm{e},j}$ of being erased. In that case, however, we can simply set p_{e} to be the worst case over $\{p_{\mathrm{e},j}\}$ and consider this more stringent model instead.

In contrast to the weak error model case, there does not exist any code, which we can use for our subroutine code, that perfectly corrects the errors at hand. Instead, there exist pretty good codes that correct the error unless too many qudits are erased. The quantum capacity of the erasure channel, with erasure probability p_{e}, has been determined to be $1 - 2p_{\mathrm{e}}$ [49]. We can choose n_{C}, the number of computational qudits, to grow with n. When n is large, since the number of qudits we want to encode is only one qudit and is much smaller than that allowed by the capacity [which is $(1-2p_{\mathrm{e}})n_{\mathrm{C}}$], the error probability would vanish exponentially in n_{C}.

We choose the number of computational qudits to scale sub-linearly in the total number of qudits, that is, $n_{\mathrm{C}} = n^\gamma$, where $\gamma \in (0, 1)$ does not depend on n and can be chosen to be very small. We use any subroutine code $(\mathcal{E}, \mathcal{D})$ that encodes one qudit into n^γ computational qudits, with the property that it has a decoding error $O(e^{-x_d \cdot n^\gamma})$ for some $x_d > 0$ that may depend on d. By a random coding argument, one can show that there exists a stabilizer code satisfying our requirement (see, e.g., Ref. [50]), although its explicit form is not given. Recently, progress in error correcting codes also showed that quantum polar codes [51,52] and Reed-Muller codes [53] have the desired property. Notice that the requirement of the $O(e^{-x_d \cdot n^\gamma})$ scaling is chosen for convenience of analyzing the error, and it can be further relaxed in practice.

Meanwhile, the model is now too noisy for the highly coupled reference frame state used in the weak error model to be effective. Instead, we prepare the reference frame register to have only two qudits per block, namely $2s_{\mathrm{R}} = n_{\mathrm{R}}$, with the reference frame on each two qudits block being the maximally entangled state. While like in the weak error model, this choice means that one erasure error only destroys one reference frame block, now there are far more blocks, which reduces the chance of them all being erased. The price to pay for this more error resilient setup, is less precise reference frames, due the smaller dimension of each block.

As long as there are still order n non erased blocks left we can achieve high performance, which happens with very high probability. Indeed, the overall error, Eq. (11), can be recast into the form:

$$\mathcal{C} = \sum_{k=0}^{n} p_{\mathrm{e}}^k (1-p_{\mathrm{e}})^{n-k} \sum_{\mathbf{s}:|\mathbf{s}|=k} (\mathcal{C}_{\mathrm{e}})_{\mathbf{s}}.$$

(12)

Then the number of surviving blocks follows a binomial distribution with mean $n(1-p_{\mathrm{e}})$, which is sufficient for a high accuracy decoding. This reasoning leads us to the following upper bound on the decoding error under the i.i.d. error model Eq. (12). For any $\alpha > 0$, there exists $n_\alpha > 0$ such that

$$\epsilon_{\mathrm{cov}} \leqslant \left(\frac{36(d^2 - d + 32)d^{\frac{d^2-d+2}{2}}}{(1-2p_{\mathrm{e}})^2 \prod_{j=1}^{d}(j-1)!} \right) \left(\frac{1}{(1-2p_{\mathrm{e}})n} \right)^{1-\alpha}$$

(13)

for all $n \geqslant n_\alpha$. Details of the proof can be found in Appendix C.

The error of our protocol scales almost as $1/n$, instead of $1/n^2$ in the previous case. This is a result of the (stronger) i.i.d. noise. In fact, the error scaling of our protocol is still optimal for this error model. Analogously to the weak error model, we can prove that the $1/n$ scaling is also optimal for the i.i.d. error
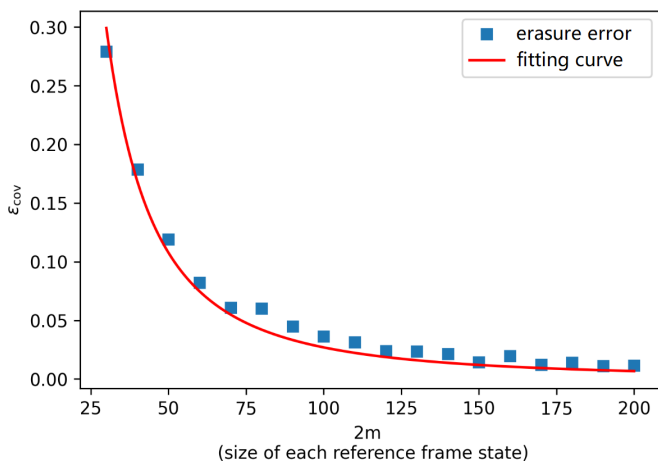
FIG. 3. Weak erasure error with "five-qubit code". Under the weak erasure model, the decoding error of our protocol (with the five-qubit code [49] as the subroutine noncovariant code and the generalized sine states as the reference frame) is plotted as a function of $2m$, the number of qubits contained in each block of the reference frame register. The blue dots correspond to the evaluated values of the decoding error and the red line corresponds to the fitting curve $\frac{269.2}{(2m)^2}$.

model. In particular, we prove that all covariant QECC satisfy the following bound under the i.i.d. error model:

$$\epsilon_{\text{cov}} \geqslant \frac{p_e}{64n(1 - p_e)}. \tag{14}$$

In the above analysis, we focused on erasure error models, as they allow us to derive analytical bounds on our protocol's performance. Nevertheless, our protocol also performs well under other commonly encountered error types, as we show in the following.

Numerical simulation for generic error models. In addition to the above bounds, we also conduct numerical experiments to evaluate the performance of our protocol. As a working example, we use the 'five-qubit code' [54] (five computational qubits; one logical qubit) as the subroutine encoding and decoding pair. We consider not only erasure errors but also other common error types like dephasing errors and depolarizing errors.

We first consider the weak erasure error model and use a group of identical generalized sine states as the reference frame register. We assume that at least one reference frame state survives the erasure error, which consists of $2m$ qubits. (Recall that the number of qubits is always even since we use entangled reference frame states.) As shown in Fig. 3, the decoding error (in the leading order) is $\frac{269.2}{(2m)^2}$, matching the prediction of our theory. We now consider errors beyond the erasure models. We first consider the case where all reference frame qubits are subject to i.i.d. depolarizing/dephasing error with probability $p = 0.2$. In this case, our reference frame register consists of $n_R/2$ maximally entangled states. Each of the reference frame states are measured individually for an estimate of the reference frame parameters. We assume at most one out of the five computational qubits is subject to error (corresponding to an error rate of 0.2), so that the logical qubit can always be recovered.
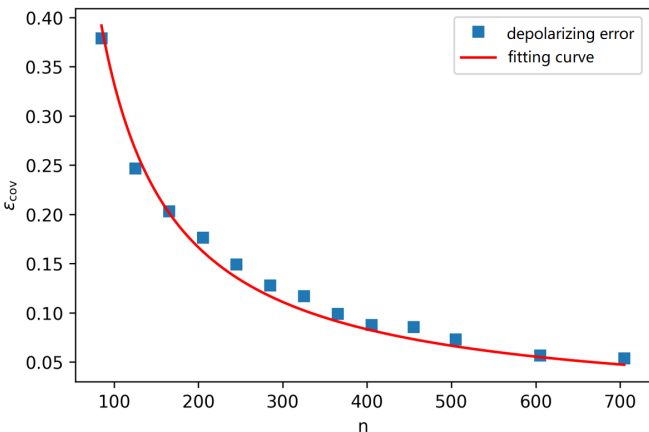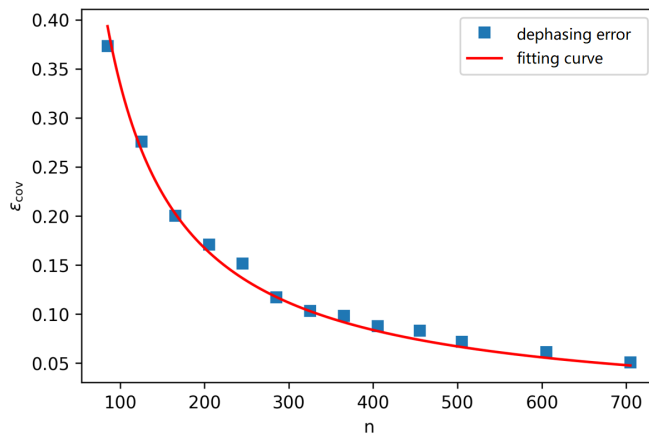

FIG. 4. i.i.d. dephasing/depolarizing with $p = 0.2$ and "five-qubit code". The decoding error for i.i.d. dephasing/depolarizing error of the our covariant code using "five-qubit code" as the subroutine encoding and $n_R/2$ maximally entangled states as the reference frame is plotted as a function of n, the number of total physical qubits. The blue dots correspond to the evaluated values of the decoding error and the red lines correspond to the fitting curves: $32.43/n$ for the dephasing error and $33.30/n$ for the depolarizing error.

Notice that, unlike the erasure error, depolarizing and dephasing errors in the reference frame register are not flagged (i.e., we do not know whether a reference frame state is subject to an error or not). In order to mitigate the effect of erroneous reference frames, we adopt a majority vote algorithm that abandons estimates in case they are likely to be faulty. With this method, for the i.i.d. depolarizing/dephasing model, we are able to achieve decoding errors (in the leading order) $\frac{33.30}{n}$ for the depolarizing error and $\frac{32.43}{n}$ for the dephasing error, with $n(:= n_R + 5)$ being the total number of physical qubits employed in our construction (see Fig. 4). The $1/n$ error scaling, observed for both the i.i.d. depolarizing error and the i.i.d. dephasing error, coincides with the lower bound Eq. (13) which we proved for the i.i.d. erasure error model. It suggests that the error scaling of our protocol is consistent in generic error models. We have also tested our protocol under variants of the above error models: We consider models where one fifth reference frame qubits go through the completely depolarizing/dephasing error, and the $1/n$ error scaling has also been observed (see Table I for a summary).

TABLE I. Performance of our construction under dephasing and depolarizing errors. For variants of the i.i.d. dephasing/depolarizing error model, the decoding errors of our construction (using "five-qubit code" as the subroutine noncovariant code) as functions of the code size are numerically evaluated. In the variant models, a randomly chosen subset of the qubits are completely depolarized/dephased, while others remain unaffected by noise. Numerical simulation shows that in models considered here the decoding errors vanish with the $1/n$ scaling.

n (total qubits)	75	135	205	305
i.i.d. dephasing ($p = 0.20$)	4.28×10^{-1}	2.58×10^{-1}	1.71×10^{-1}	1.18×10^{-1}
i.i.d. depolarizing ($p = 0.20$)	4.29×10^{-1}	2.62×10^{-1}	1.77×10^{-1}	1.27×10^{-1}
20% qubits dephased	3.88×10^{-1}	2.38×10^{-1}	1.74×10^{-1}	1.20×10^{-1}
20% qubits depolarized	4.60×10^{-1}	3.15×10^{-1}	2.18×10^{-1}	1.53×10^{-1}
n (total qubits)	405	505	605	705
i.i.d. dephasing ($p = 0.20$)	8.79×10^{-2}	7.19×10^{-2}	6.02×10^{-2}	5.09×10^{-2}
i.i.d. depolarizing ($p = 0.20$)	8.76×10^{-2}	7.32×10^{-2}	5.83×10^{-2}	5.30×10^{-2}
20% qubits dephased	8.65×10^{-2}	7.03×10^{-2}	5.94×10^{-2}	5.34×10^{-2}
20% qubits depolarized	1.21×10^{-1}	9.17×10^{-2}	7.52×10^{-2}	6.37×10^{-2}

These results show that the application of our protocol is not limited to erasure errors. Instead, it has good potential to work well under generic error types. Our code is available at Ref. [55], and more details can be found in Appendix E.

Resource requirements and implementation. We now analyze some of the memory and time resources needed to implement our QECC which are required beyond those of the subroutine code.

Our reference frames are constructed on n_R qudits constituting a d^{n_R} dimensional Hilbert space. However, the reference frame states only have support on a space $d_R :=$ $\dim(\mathsf{Span}\{\mathcal{U}_R(\Psi)\}_{U \in \mathsf{SU}(d)})$, which we prove to be upper bounded by

$$d_R \leqslant \left(\frac{n_R}{2} + 1\right)^{(d^2-1)}. \tag{15}$$

Therefore we can compress the reference frame state to a system of much smaller dimension, which reduces exponentially the cost of quantum memory. At any time when the reference frame state is idling, it can be compressed by a covariant isometry into a much smaller system. The cost can be further reduced at the price of a small recovery error using the compression protocols in Refs. [56–58].

We now turn our attention to the implementation of the covariant encoding and decoding: At first sight, our encoding and decoding may appear to necessitate sampling from the Haar measure on $\mathsf{SU}(d)$ in order to be implemented [see Eqs. (3) and (4)]. From a computational complexity standpoint, this would be a problem, since to implement a Haar-random unitary one needs an exponential number of two-qubit gates and random bits [59]. Fortunately, our encoding and decoding do not require sampling from the Haar measure, but rather only from a distribution which agrees with it up to the tth moment, for appropriate t. Such equivalent distributions are known as unitary t-designs [37–44], and are more efficient to implement.

Specifically, let U be a unitary representation on $\mathbb{C}^{d \times d}$ and $P_{t,t}(U)$ be a matrix whose entries are polynomials of order t in the coefficients of U and of order t in the coefficients of U^*, and let $\mathbb{E}_{U \sim \nu}[f(U)]$ be the expectation value of a function f according to measure ν. We then say that $\mathbb{E}_{U \sim \nu_{\text{Haar}}}[P_{t,t}(U)]$,

where ν_{Haar} denotes the Haar measure, *admits a unitary t-design.*

Our result is that the encoder $\mathcal{E}_{\text{cov}}(X_L)$ and decoder $\mathcal{D}_{\text{cov}}(X_C)$ admit unitary $(n_C + n_R/2 + 1)$-designs for all $X_L \in \mathsf{Lin}(\mathcal{H}_L)$, $X_C \in \mathsf{Lin}(\mathcal{H}_C^{\otimes n})$. See Appendix D 1 for proof. Ref. [43] devises and quantifies a method to approximate unitary t-designs: to approximate our encoder and decoder up to an error ϵ_{cov} in diamond norm, it is enough to use k two-qubit gates, randomly drawn from the Haar measure on $\mathsf{SU}(4)$. Here

$$k = 510\,000 \log_2(d)(n_C + n_R/2 + 1)^{9.1}$$
$$\times (2\log_2(d)(n_C + n_R/2 + 1) + 1 + \log_2(1/\epsilon_{\text{cov}})), \tag{16}$$

which scales polynomially in both n_C and n_R [recall lower bounds on ϵ_{cov}, in the weak and i.i.d. error models, Eqs. (10) and (14)].

Since sampling a polynomial number of times from $\mathsf{SU}(4)$ can be performed efficiently, both the encoder and the decoder can be efficiently implemented in n_C and n_R so long as the reference frame state Ψ and measurement $\mathcal{M}_{\hat{U}}$ can be efficiently constructed.

Efficient error syndrome extraction. In quantum error correction, we often need to measure error syndromes after applying quantum gates. It might appear that a full decoding is needed to achieve this in our QECC, which, despite the efficiency analysis in the previous part, could have non-negligible effects on the performance of the computation. However, this is not true for the erasure error models we considered. Instead, one can keep track of the error in a very simple way and correct them all together at a suitable time.

To see this, first notice that erasure errors are "flagged", meaning that the qudit is mapped from its original Hilbert space $\mathcal{H}$ to an orthogonal Hilbert space $\mathcal{H}_{\text{err}}$. Then, the error syndrome can be obtained by a binary projective measurement consisting of projectors on $\mathcal{H}$ and $\mathcal{H}_{\text{err}}$. This measurement commutes with the (covariant) implementation of logical operations, and thus to implement a sequence of logical gates we can adapt the following procedure.

(1) Apply $\mathcal{E}_{\text{cov}}$.

(2) Create a (classical) bit string s, initiated as a null vector, to record erasure errors.

(3) For $i = 1, \dots, k$:

(a) apply $\mathcal{V}_i$,

(b) measure error syndromes, and

(c) add the measured error locations to s.

(d) If |s| exceeds the maximal number of tolerable local errors, do $\mathcal{D}_{cov}$ followed by $\mathcal{E}_{cov}$.

(4) Apply $\mathcal{D}_{cov}$.

Therefore a full physical decoding is done only if there is too many local errors (which is determined by the subroutine code) rather than after each logical gate, making our QECC much more efficient than it appears. We remark that how this feature generalizes to other error models remains an open question to be investigated in the future.

Towards fault-tolerant QECCs. To achieve practical fault tolerance, one must show that each step of the QEC procedure can be performed fault-tolerantly. For our QECC construction, the logical gate implementation is transversal and thus fault-tolerant thanks to covariance. It remains to be shown that the encoder and the decoder are both (a) computationally efficient to implement and (b) can be implemented fault tolerantly. For (a), we have provided covariant encoders and decoders which can be implemented via a polynomially in n (total number of physical qudits) number of two-qubit gates, while still maintaining the optimal scaling of the decoding error. Regarding (b), making the encoder and the decoder fault-tolerant amounts to finding a fault-tolerant implementation of the $\mathsf{SU}(d)$ twirling in Eqs. (3) and (4). We remark that it is unlikely that unitary designs are the solution: any unitary t-design D is an ϵ-net on the parameter space of $\mathsf{SU}(d)$, with ϵ vanishing in t. For large t, the ϵ-net becomes dense enough that, unless the commutant of $\{U \otimes U^* \,|\, U \in \mathsf{D}\}$ has dimension larger than two, D forms a universal gate set (see also Ref. [60, remark 4]). In this case, it follows that to implement the unitary design fault tolerantly we need to implement a universal gate set fault tolerantly, leading to a contradiction.

Nevertheless, unitary designs are not a must for implementing the twirling in Eqs. (3) and (4). One can, for example, use extrapolation of different noisy channels as an alternative approach. Similar ideas have resulted in error mitigation [61,62], which effectively deals with errors in near-term devices.

Another possibility, is to partition the logical space into code blocks and apply our encoding to each block individually (call this level 1 encoding). One then implements the gates corresponding to the computation such that any one gate has support on either one or two code blocks. When one wishes to implement a gate across two code blocks, these code blocks are encoded once more using our encoding scheme (call this level 2 encoding); the unitary is then implemented and we decode back to the level 1 encoding. This way, one can increase the logical code size by adding more code blocks (each one of constant size) while only applying gates on encoded states. The quantum reference frames for the level 2 encoding can be repurposed after ever level 2 decoding for level 2 encodings over other code blocks. Using this scheme, there is a much better overall scaling of the resources involved since we no longer require $\mathsf{SU}(d)$ twirling over the entire logical space. These possibilities will be explored more in future research. In summary, our covariant QECC construction does not immediately lead to a satisfactory approach of achieving full fault tolerance, but it opens up a new route that is worth further exploration.

III. DISCUSSION

Starting from an arbitrary error correcting code, we have shown how to use quantum reference frames to construct a new code for which all gates are covariant and transversal, up to a small error in the decoding. In the absence of noise, the error in the decoding is ultimately a consequence of the inability to perfectly determine the orientation of the reference frames via measurement.

Our results generalize the setting in Ref. [5] from d-dimensional reference frame representations of $\mathsf{U}(1)$ to d-dimensional reference frame representations of $\mathsf{SU}(d)$. This construction is important primarily since once the logical states have been encoded, it allows for a universal set of gates to be applied transversally, which in turn can allow for errors to be more easily detected and corrected during a computation. Furthermore, the construction also allows any quantum error correcting code to be converted into a covariant code. Previously it was not known if this is possible with a finite dimensional physical space. In the setting of the weak error model and $\mathsf{SU}(d)$ symmetry, we establish novel upper bounds on the decoding error which are tight in the scaling with n (number of physical subsystems). Indeed, previous upper bounds in this context [6,63] are not tight with n. Furthermore, in this context, we also improved previous results [47,48] on the lower bounds of the decoding error. We have also analysed the i.i.d. erasure model, which is more realistic in the context of computation. Here we have produced upper and lower bounds for the decoding error which are tight in the scaling with n. We argued that our model for a QECC should be able to correct errors beyond those of the local erasure type. Numerical simulations of the performance of our construction for practical error models including the depolarizing error and the dephasing error bears this out.

Our results also feature applications beyond correcting errors in quantum computation. Two party communication with misaligned reference frames and lossy quantum channels [34] is a clear example, but the range of potential applications is much broader. In particular, we designed and introduced generalized sine states and used them as optimal reference frames in the weak error model. A characteristic novel feature of the generalized sine states is that they can achieve the Heisenberg limit in estimating a completely unknown unitary gate, making it promising to apply them in the emergent field of multiparameter quantum metrology [64–68] and, more generally, in quantum information processing. Remarkably, these states have recently been used as the key to solving the long-standing optimal programming problem of unitary quantum gates [69].

From a more fundamental physics perspective, our results provide new insight into the study of quantum gravity. In particular, there is a conjectured duality between a theory of gravity in anti-de Sitter (AdS) space and a conformal quantum field theory (CFT) on its boundary, known as the AdS-CFT correspondence. The duality is mediated via a quantum error correcting code, where the encoding channel maps low-energy states in the anti-de Sitter space (known as the "bulk") to states on the conformal field (known as the "boundary") [70]. The implication of the Eastin-Knill theorem, discussed in the introduction, to the AdS-CFT duality is that global continuous

symmetries in the bulk cannot map onto local symmetries on the boundary (see [71,72] for details). This is often interpreted as a problem for the duality since initially such symmetries were believed to exist. It is thus interesting to quantify, as a function of physical quantities such as energy or dimension, how local symmetries of the bulk can approximate local symmetries in the boundary. This program was initiated in Refs. [5,6,73] and is the subject of ongoing investigation (see, e.g., Ref. [74]). Our results quantify the extent to which the approximate duality for global $\mathsf{SU}(d)$ symmetries could exist.

Our numerical results suggest that our reference frame based QECCs work well for generic nondetectable error models. Understanding and analytically bounding the decoding error, on the other hand, requires further investigation. It would be an interesting future direction of research to rigorously analyze the performance of our QECCs in generic error models and compare to other existing constructions [63]. The dephasing error model, for example, was studied in the U(1) case in Ref. [5] and shown to be correctable. This should also be true in the current $\mathsf{SU}(d)$ case. These studies will also help pave the way to proving robustness of our results when optimized for error models found in experiments.

Finally, since our protocols only involve classical correlations between the computational system and the reference frame system, this setup is naturally suited to implementing these two physical subsystems on different quantum architectures to harness specific advantages of each platform. For example, one may wish to implement the quantum reference frames on a phonic platform using Gaussian quantum computing [75], while using superconducting qubits for the computational space.

APPENDIX A: CONSTRUCTION OF COVARIANT ERROR CORRECTING CODES

1. Preliminary

We use the notation $\mathcal{H}_x$ ($x =$ R, P, L) to denote the Hilbert space of the reference frame register, the computational register, or the logical register. In particular, $\mathcal{H}$ refers to a d-dimensional Hilbert space. For a generic Hilbert space $\mathcal{K}$ and a pure state $|\psi\rangle \in \mathcal{K}$, we will use the notation $\psi := |\psi\rangle\langle\psi|$ to denote the projector on the one-dimensional subspace spanned by $|\psi\rangle$. The set of quantum states will be denoted by $\mathsf{St}(\mathcal{K})$. In this work we will focus on finite dimensional quantum systems, with $\dim(\mathcal{K}) < \infty$.

A quantum process transforming an input system into a (possibly different) output system is called a *quantum channel*. A quantum channel transforming an input system with Hilbert space $\mathcal{K}_{\mathrm{in}}$ into an output system with Hilbert space $\mathcal{K}_{\mathrm{out}}$ is a completely positive trace-preserving map $\mathcal{N} : \mathsf{Lin}(\mathcal{H}_{\mathrm{in}}) \to \mathsf{Lin}(\mathcal{H}_{\mathrm{out}})$, where $\mathsf{Lin}(\mathcal{K})$ denotes the space of linear operators on $\mathcal{K}$. For pure channels with a single Kraus operator U, i.e., unitaries or isometries, we use the shorthand $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$.

We will use the big-Ω notation, the big-O notation, and the big-Θ notation to characterize the asymptotic behavior of functions. For a function $f(n)$, we write $f(n) = \Omega(g(n))$ if there exists a constant $c_1 > 0$ so that $f(n) \geqslant c_1 g(n)$ for large enough n, $f(n) = O(g(n))$ if there exists a constant $c_2 > 0$ so that $|f(n)| \leqslant c_2 g(n)$ for large enough n, and $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$.

We will make frequent use of a couple of basic concepts in representation theory of the special unitary group $\mathsf{SU}(d)$, which can be found in standard textbooks, e.g., Ref. [76]. We denote the Young diagrams by a vector[3] $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots)$ with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$ and by $U_{\boldsymbol{\lambda}}$ the irreducible representation of $\mathsf{SU}(d)$ characterized by the Young diagram $\boldsymbol{\lambda}$. We denote the collection of all Young diagrams with n boxes and at most d rows by Y_n (since the dimension d is fixed throughout the paper, we often omit it). In particular, we define $\boldsymbol{e}_i$ to be the vector whose i-th entry is one and other entries are zero. By definition, $\boldsymbol{e}_1$ corresponds to a legitimate Young diagram whose associated representation is the standard one, and we use the abbreviation $U := U_{\boldsymbol{e}_1}$. We will use the double-ket notation $|A\rangle\rangle := \sum_{n,m} \langle n|A|m\rangle |n\rangle|m\rangle$ ($\{|n\rangle\}$ being an orthonormal basis) with A being a matrix. With the double-ket notation, we denote by $|\Phi^+_{U,\boldsymbol{\lambda}}\rangle := |U_{\boldsymbol{\lambda}}\rangle\rangle/\sqrt{d_{\boldsymbol{\lambda}}}$ the maximally entangled state associated to $U_{\boldsymbol{\lambda}}$, with $d_{\boldsymbol{\lambda}} = \mathrm{Tr}\, I_{\boldsymbol{\lambda}}$ being the dimension of the irreducible subspace. By this notation, $|\Phi^+_U\rangle$ refers to a maximally entangled state in $\mathcal{H} \otimes \mathcal{H}$, rotated by a local unitary U. Unless otherwise specified, dU or sometimes $d\hat{U}$ denotes the Haar measure.

Next we introduce a couple of measures of error and faithfulness that are used in this work. Consider two channels $\mathcal{A}, \mathcal{B} : \mathsf{Lin}(\mathcal{K}_{\mathrm{in}}) \to \mathsf{Lin}(\mathcal{K}_{\mathrm{out}})$. For instance, $\mathcal{A}$ can be a desired quantum gate and $\mathcal{B}$ is an approximation of $\mathcal{A}$. Therefore we often need measures to quantify their distance (i.e. measures of simulation error) and similarity (i.e. measures of simulation faithfulness).

The first measure of faithfulness, defined intuitively as the minimum fidelity between the outputs for the same input, is the *worst-case input (or minimum) fidelity*:

$$F_{\mathrm{wc}}(\mathcal{A}, \mathcal{B}) := \inf_{\rho \in \mathsf{St}(\mathcal{K}_{\mathrm{in}} \otimes \mathcal{K}')} F((\mathcal{A} \otimes \mathcal{I}_{\mathcal{K}'})(\rho), (\mathcal{B} \otimes \mathcal{I}_{\mathcal{K}'})(\rho)),$$

$$(\mathrm{A1})$$

---

[3]For vectors, we use the bond font.

where $F(\rho, \sigma) := (\mathrm{Tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}})^2$ is the Uhlmann fidelity defined for two arbitrary quantum states ρ and σ and $\mathcal{K}'$ is an arbitrary reference Hilbert space. Later we give a semidefinite program to compute $F_{wc}(\mathcal{A}, \mathcal{B})$ (see lemma 6). Nonetheless, the worst-case fidelity is often difficult to work with, and thus it is more common to consider another measure of faithfulness, known as the *entanglement fidelity* [77]:

$$F_{\mathrm{ent}}(\mathcal{A}, \mathcal{B}) := F((\mathcal{A} \otimes \mathcal{I}_{\mathcal{K}'})(\Phi^+), (\mathcal{B} \otimes \mathcal{I}_{\mathcal{K}'})(\Phi^+)), \quad (A2)$$

where $\Phi^+ := |\Phi^+\rangle\langle\Phi^+|$ with $|\Phi^+\rangle \in \mathcal{K}_{\mathrm{in}} \otimes \mathcal{K}'$ being the maximally entangled state of the system and a reference. It is straightforward by definition that $F_{wc} \leqslant F_{\mathrm{ent}}$.

On the other hand, one can also quantify the performance of simulation by error measures. Similarly as for faithfulness, there are also two commonly considered measures of error, defined with respect to the worst-case input and the maximally entangled state. The first is the worst-case input error, or the *diamond norm error*, defined as

$$\epsilon_{wc}(\mathcal{A}, \mathcal{B}) := \tfrac{1}{2}\|\mathcal{A} - \mathcal{B}\|_\diamond \qquad (A3)$$

$$= \tfrac{1}{2} \sup_{\rho \in \mathsf{St}(\mathcal{K}_{\mathrm{in}} \otimes \mathcal{K}')} \|(\mathcal{A} \otimes \mathcal{I}_{\mathcal{K}'})(\rho) - (\mathcal{B} \otimes \mathcal{I}_{\mathcal{K}'})(\rho)\|_1,$$
$$(A4)$$

where $\|\cdot\|_1$ is the trace norm and $\|\cdot\|_\diamond$ denotes the *diamond norm* [78]. Correspondingly, we have the *entanglement error*:

$$\epsilon_{\mathrm{ent}}(\mathcal{A}, \mathcal{B}) := \tfrac{1}{2}\|(\mathcal{A} \otimes \mathcal{I}_{\mathcal{K}'})(\Phi^+) - (\mathcal{B} \otimes \mathcal{I}_{\mathcal{K}'})(\Phi^+)\|_1. \quad (A5)$$

Again, by definition, we have $\epsilon_{\mathrm{ent}} \leqslant \epsilon_{wc}$.

The above faithfulness measures and error measures are not independent [79]. For instance, the well-known Fuchs-van de Graaf inequalities [80], which relate fidelities to distances as

$$1 - \sqrt{F(\rho, \sigma)} \leqslant \tfrac{1}{2}\|\rho - \sigma\|_1 \leqslant \sqrt{1 - F(\rho, \sigma)} \quad \forall \rho, \sigma$$
$$(A6)$$

can be readily employed. Besides, the worst-case input error can be bounded using the following inequality

$$\epsilon_{wc}(\mathcal{A}, \mathcal{B}) \leqslant d_S \cdot \epsilon_{\mathrm{ent}}(\mathcal{A}, \mathcal{B}), \qquad (A7)$$

where d_S is the dimension of the system [81, exercise 3.6].

2. Quantum reference frames

In this section, we introduce quantum reference frames, which will be the key ingredient of our construction of covariant error correcting codes. We will consider entangled reference frames, as they offer superior performance [17,35,36]. A quantum reference frame is essentially an entangled state $|\psi'\rangle_{AB}$, constructed on $2m$ qudits and shared by two parties Alice and Bob. Assume that the reference frames of Alice and Bob are misaligned up to a rotation $U \in \mathsf{SU}(d)$: if Bob prepares one of his qudits in, say, a state $|k\rangle$ in the computational basis, it would actually be a rotated version of the corresponding state on Alice's side, i.e., $|k\rangle_B = U|k\rangle_A$.

To align their reference frames, Bob sends his part of the state through a distortion-free channel to Bob. Now, Alice has full disposal of the bipartite state $|\psi'\rangle_{AA'} = (U^{\otimes m} \otimes I^{\otimes m})|\psi\rangle$,

where $|\psi\rangle$ is the state when there is no reference frame misalignment. By making a suitable measurement, Alice obtains information on the misalignment U to a good precision.

In this work, we will mainly use a quantum reference frame in a single-party but mathematically equivalent scenario. The role of quantum reference frames in our construction of covariant error correcting codes is to store an unknown unitary rotation. By the Schur-Weyl duality [76], we have the following decomposition of the product Hilbert space into the direct sum of irreducible subspaces:

$$\mathcal{H}^{\otimes m} \simeq \bigoplus_{\lambda \in \mathsf{Y}_m}(\mathcal{H}_\lambda \otimes \mathcal{M}_\lambda). \qquad (A8)$$

Here Y_m is the collection of Young diagrams of m boxes, $\mathcal{H}_\lambda$ and $\mathcal{M}_\lambda$ denote the irreducible subspace and the multiplicity subspace associated to the Young diagram λ, respectively. Notice that $\mathcal{M}_\lambda$ also depends on m but this dependence is not important in this work. The notation "$\simeq$" means that the two spaces are equivalent up to a change of basis, which is called the Schur transform and can be implemented efficiently on a quantum computer [82,83].

In our work, it is enough to consider (quantum) reference frame states of the form [36]

$$|\psi\rangle = \bigoplus_{\lambda \in \mathsf{Y}_m}\sqrt{q_\lambda}|\Phi_\lambda^+\rangle \otimes |\Phi_{m_\lambda}^+\rangle, \qquad (A9)$$

where $|\Phi_\lambda^+\rangle$ and $|\Phi_{m_\lambda}^+\rangle$ denote the maximally entangled states of the irreducible subspace and the multiplicity subspace, respectively, and $\{q_\lambda\}$ is a probability distribution that uniquely characterizes the reference frame state. Here the state is built on the coupled basis of $2m$ qudit systems, half of which serve as a reference. Explicitly, it lies within the Hilbert space $\bigoplus_\lambda \mathcal{H}_\lambda \otimes \mathcal{H}'_\lambda \otimes \mathcal{M}_\lambda \otimes \mathcal{M}'_\lambda \subset \mathcal{H}^{\otimes m} \otimes \mathcal{H}^{\otimes m}$, where $\mathcal{H}_\lambda (\simeq \mathcal{H}'_\lambda)$ is the representation subspace and $\mathcal{M}_\lambda (\simeq \mathcal{M}'_\lambda)$ is the multiplicity subspace.

The reference frame state lies in a subspace of $\mathcal{H}^{\otimes m} \otimes \mathcal{H}^{\otimes m}$. To it we apply $(U \otimes I)^{\otimes m}$ on it,[4] resulting in a state of the form

$$|\psi_U\rangle = \bigoplus_\lambda \sqrt{q_\lambda}|\Phi_{U,\lambda}^+\rangle \otimes |\Phi_{m_\lambda}^+\rangle. \qquad (A10)$$

The optimal measurement to extract the information of U from the above state is the covariant POVM $\{d\hat{U}, |\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|\}$ [36], where $d\hat{U}$ is the Haar measure and $|\eta_{\hat{U}}\rangle$ is the following vector:

$$|\eta_{\hat{U}}\rangle := \bigoplus_\lambda d_\lambda |\Phi_{\hat{U},\lambda}^+\rangle \otimes |\Phi_{m_\lambda}^+\rangle, \qquad (A11)$$

where d_λ denotes the dimension of the irreducible subspace characterized by λ. Denoting by $\chi_{U,\lambda} := \mathrm{Tr}[U_\lambda]$ the characters of $\mathsf{SU}(d)$, the probability density function of getting the

---

[4]A covariant code in the usual sense should have realization $U^{\otimes 2m}$ on it. Here instead we realize it as $(U \otimes I)^{\otimes m}$, which is still both mathematically rigorous ($U \otimes I$ is still a representation of U) and good for the physical implementation (just run fewer operations).

outcome $\hat{U}$ when the actual gate is U can be expressed as

$$p(\hat{U}|U) = \left| \sum_\lambda \sqrt{q_\lambda} \chi_{U\hat{U}^{-1},\lambda} \right|^2. \qquad (A12)$$

This distribution is nice because it is conjugate invariant, i.e., $p(WUV|WV) = p(U|I) = p(U^\dagger|I)$ for any W, V, and U. The above defines a series of estimation strategies, each specified by a choice of $\{q_\lambda\}$. Later, we will make specific choices that fulfill the goal of constructing covariant codes.

In the error correction scenario, we have to deal with errors occurring on the reference frames. To this purpose, we divide the n_R qudits assigned to the reference frame register into $s_R > 1$ groups of identical systems, i.e., $\mathcal{H}_R = \mathcal{H}_{R,1} \otimes \mathcal{H}_{R,2} \otimes \cdots \otimes \mathcal{H}_{R,s_R}$ and construct an individual reference frame state of the form in Eq. (A9) on each of them. We assume that the error occurring on each of the reference frames is detectable, i.e., it takes the state on $\mathcal{H}_{R,k}$ ($\forall k$) to an error space $\mathcal{H}_{\mathrm{err},k}$ that is orthogonal to $\mathcal{H}_{R,k}$. This is the case, for instance, for erasure errors.

For detectable errors, we will use the reference frames in the following way.

(1) We prepare all the reference frame registers in the same state Eq. (A9) and denote the overall state by $\Psi := \psi^{\otimes s_R}$. Here, since each ψ is constructed on $2m$ qudits, we have $2m s_R = n_R$.

(2) For each copy of ψ, we perform a binary POVM to detect whether an error has happened. For each label k, the POVM is defined by projections on the error space $\mathcal{H}_{R,\mathrm{err},k}$ and the original space $\mathcal{H}_{R,k}$.

(3) We discard those erroneous reference frames and measure *jointly* the remaining ones with the measurement defined in Eq. (A11).

Denote by $\mathcal{H}_{\mathrm{err},R} := \mathcal{H}_{\mathrm{err},R,1} \otimes \cdots \otimes \mathcal{H}_{\mathrm{err},R,s_R}$ the total error space. The POVM element of our error-tolerant measurement is $\{|\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|_\mathbf{s} \otimes P_{\mathrm{err},\mathbf{s}^c}, d\hat{U}\}_{\mathbf{s}\subset[s_R]}$, where $\{|\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|_\mathbf{s}, d\hat{U}\}$ corresponds to a joint measurement on $\bigotimes_{k\in\mathbf{s}} \mathcal{H}_{R,k}$ with $|\eta_{\hat{U}}\rangle$ defined by Eq. (A11), $P_{\mathrm{err},\mathbf{s}^c} := \bigotimes_{k\in\mathbf{s}^c} P_{\mathrm{err},k}$ is the projection on the error spaces of all $k \notin \mathbf{s}$, and $[l]$ for $l \in \mathbb{N}$ denotes the set $\{1, 2, \ldots, l\}$. Alternatively, the measurement is characterized by a linear map that yields a probability density function $\mathcal{M}_{\hat{U}} : \mathsf{St}(\mathcal{H}_R \otimes \mathcal{H}_{\mathrm{err},R}) \to \mathbb{R}$:

$$\mathcal{M}_{\hat{U}}(\cdot) := \sum_{\mathbf{s}\subset[s_R]} \mathrm{Tr}\left[(\cdot)|\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|_\mathbf{s} \otimes P_{\mathrm{err},\mathbf{s}^c}\right]. \qquad (A13)$$

When an error $\mathcal{C}_R$ occurs, the probability density function of the outcome $\hat{U}$ is just $\mathcal{M}_{\hat{U}} \circ \mathcal{C}_R(\Psi)$.

3. Covariant quantum error correction

Given n qudit systems in total, the goal of covariant quantum error correction is to construct a covariant code $(\mathcal{E}_{\mathrm{cov}}, \mathcal{D}_{\mathrm{cov}})$ that encodes one logical qudit, i.e., $d_L = d$ into the n qudit systems. This is to say that the encoder $\mathcal{E}_{\mathrm{cov}}$ satisfies

$$\mathcal{E}_{\mathrm{cov}} \circ \mathcal{V}_L = (\mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_n) \circ \mathcal{E}_{\mathrm{cov}} \qquad (A14)$$

for any $V \in \mathsf{SU}(d)$, where $\mathcal{V}_i$ ($i = 1, \ldots, n$) is either fixed to be $\mathcal{V}$ or fixed to be the identity channel. In this way, any logical gate can be realized *transversally* by implementing the same

gate on each of the qudit systems, which is the desideratum of fault-tolerant quantum computing.

Nevertheless, the Eastin-Knill theorem rules out any covariant, finite dimensional code that perfectly corrects local errors. One way to circumvent this restriction is to consider approximate codes that are still covariant but correct an error $\mathcal{C}$ only ϵ-well:

$$\mathcal{D}_{\mathrm{cov}} \circ \mathcal{C} \circ \mathcal{E}_{\mathrm{cov}} \approx_\epsilon \mathcal{I}_L. \qquad (A15)$$

This was first achieved for $\mathsf{U}(1)$ covariance using finite dimensional reference frames in Ref. [5] and using other techniques in Ref. [6]. The main idea of our construction is to divide the n qudit systems into two registers: the reference frame register R (consisting of n_R qudits), and the computational register C (consisting of $n_C = n - n_R$ qudits). We denote by $\mathcal{U}_C := \mathcal{U}^{\otimes n_C}$ and $\mathcal{U}_L := \mathcal{U}$ the actions of an element $U \in \mathsf{SU}(d)$ on the computational register and the logical register, and by $\mathcal{U}_R := (\mathcal{U} \otimes \mathcal{I})^{\otimes n_R/2}$ its action on the reference frame register, were recall $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$. Our covariant code is built on an arbitrary, possibly noncovariant, code $(\mathcal{E}, \mathcal{D})$ on the computational register, i.e., $\mathcal{E} : \mathcal{H}_L \to \mathcal{H}_C$. We make $\mathcal{E}$ and $\mathcal{D}$ covariant via the technique of twirling. The role of the reference frame register is to keep track of the random unitaries applied in the twirling and to keep the subroutine code functioning.

In particular, the covariant encoder is defined by Eq. (3) and the decoder is defined by Eq. (4) with $\mathcal{M}_{\hat{U}}$ the measurement from Eq. (A13).

4. General bound on the error

There are two contributing factors to errors in the covariant scheme: noise occurring between encoding and decoding, and errors in the decoding due to the finiteness of the reference frames. Here we show that our protocol is able to recover from both of these. We quantify the error by the diamond norm error Eq. (A4)

$$\epsilon_{\mathrm{cov}} := \max_{V\in\mathsf{SU}(d)} \epsilon_{\mathrm{wc}}(\mathcal{D}_{\mathrm{cov}} \circ \mathcal{C} \circ (\mathcal{V}_C \otimes \mathcal{V}_R) \circ \mathcal{E}_{\mathrm{cov}}, \mathcal{V}_L), \quad (A16)$$

where $\mathcal{C}$ is a noisy channel. Here we focus on errors that are detectable on the reference frames (see Sec. A 2) and covariant. We assume that the error can be expressed in the form

$$\mathcal{C} = \sum_j p_j \mathcal{C}_j, \qquad (A17)$$

where each $\mathcal{C}_j$ is of the product form

$$\mathcal{C}_j = \mathcal{C}_{j,C} \otimes \mathcal{C}_{j,R} \qquad (A18)$$

acting on the computational register and the reference frame register, respectively. For the subroutine code pair $(\mathcal{E}, \mathcal{D})$, let $\epsilon_{j,\mathrm{code}} = \epsilon_{\mathrm{wc}}(\mathcal{D} \circ \mathcal{C}_{j,C} \circ \mathcal{E}, \mathcal{I}_L)$ denote the worst-case error for noise operator $\mathcal{C}_j$. Then we can show

Lemma 1. For covariant error of the form Eq. (A17), the error of our construction (see protocol 1) II is upper bounded as

$$\epsilon_{\mathrm{cov}} \leqslant 9d \cdot \sum_j p_j \epsilon_j, \qquad (A19)$$
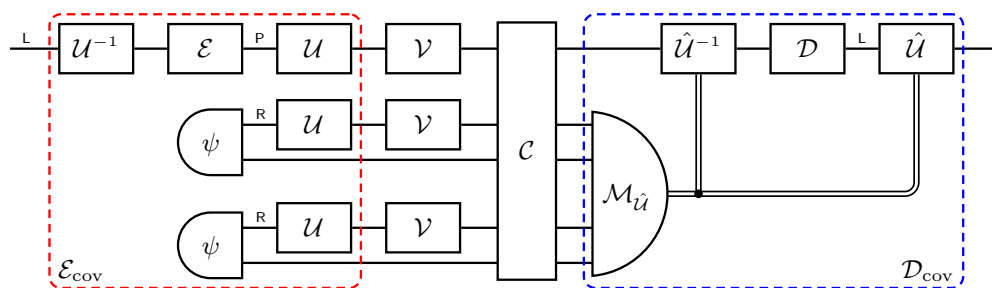
FIG. 5. Reference frame assisted covariant error correction. An arbitrary code $(\mathcal{E}, \mathcal{D})$ can be converted into a covariant code $(\mathcal{E}_{\text{cov}}, \mathcal{D}_{\text{cov}})$ following the instruction of Protocol 1. The subroutine encoder $\mathcal{E} : \mathcal{H}_{\text{L}} \rightarrow \mathcal{H}_{\text{C}}$ is made covariant via twirling, where U is a random unitary sampled from the Haar measure. A reference frame register, which is divided into multiple copies of a reference frame state ψ, is employed to keep track of the unitary. Since $\mathcal{E}_{\text{cov}}$ is covariant, any desired logical operation V can be implemented transversally. To correct error (characterized by a channel $\mathcal{C}$) or to decode the information, the reference frame register is first measured, and recovery operations are performed on the computational register depending on the measurement outcome.

where

$$\epsilon_j = \max \left\{ \epsilon_{j,\text{code}}, 1 - F_{\text{wc}} \right.$$
$$\left. \times \left(\int dU \, p_j(U|I) \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{L}}^*, \mathcal{I}_{\text{C}} \otimes \mathcal{I}_{\text{L}} \right) \right\} \quad (A20)$$

for $F_{\text{wc}}(\int dU \, p_j(U|I) \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{L}}^*, \mathcal{I}_{\text{C}} \otimes \mathcal{I}_{\text{L}}) \geqslant 3/4$; otherwise $\epsilon_j = 1$. Here

$$p_j(\hat{U}|U) := \mathcal{M}_{\hat{U}} \circ \mathcal{C}_{j,\text{R}} \circ \mathcal{U}_{\text{R}}(\Psi) \quad (A21)$$

is the probability distribution of the measurement outcome conditioned on U being applied and error $\mathcal{C}_j$ occurring.

We remark that, since we mainly focus on the small error regime, the condition $F_{\text{wc}} \geqslant 3/4$ in the above lemma is usually guaranteed.

Proof. Since the protocol $(\mathcal{E}_{\text{cov}}, \mathcal{D}_{\text{cov}})$ and the error $\mathcal{C}$ are both covariant, the protocol fares equally well for any $\mathcal{V}_{\text{L}}$, and thus we have

$$\epsilon_{\text{cov}} = \epsilon_{\text{wc}}(\mathcal{D}_{\text{cov}} \circ \mathcal{C} \circ \mathcal{E}_{\text{cov}}, \mathcal{I}_{\text{L}}). \quad (A22)$$

Substituting the expressions of $\mathcal{E}_{\text{cov}}$ [see Eq. (3)] and $\mathcal{D}_{\text{cov}}$ [see Eq. (4)] as well as $\mathcal{C} = \sum_j p_j \mathcal{C}_j$ into the definition, we obtain

$$\epsilon_{\text{cov}} \leqslant \sum_j p_j \epsilon_j \quad (A23)$$

$$\epsilon_j := \epsilon_{\text{wc}} \left(\int dU \int d\hat{U} \, (\hat{\mathcal{U}}_{\text{L}} \circ \mathcal{D} \circ \hat{\mathcal{U}}_{\text{C}}^{-1} \otimes \mathcal{M}_{\hat{U}}) \circ \mathcal{C}_j \circ (\mathcal{U}_{\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1} \otimes \mathcal{U}_{\text{R}}(\Psi)), \mathcal{I}_{\text{L}} \right), \quad (A24)$$

having used the joint convexity of the diamond norm error. Since each $\mathcal{C}_j$ is decomposed as $\mathcal{C}_{j,\text{C}} \otimes \mathcal{C}_{j,\text{R}}$, we find

$$\epsilon_j = \epsilon_{\text{wc}} \left(\int dU \int d\hat{U} \, p_j(\hat{U}|U) \hat{\mathcal{U}}_{\text{L}} \circ \mathcal{D} \circ \hat{\mathcal{U}}_{\text{C}}^{-1} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{U}_{\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1}, \mathcal{I}_{\text{L}} \right), \quad (A25)$$

where $p_j(\hat{U}|U) := \mathcal{M}_{\hat{U}} \circ \mathcal{C}_{j,\text{R}} \circ \mathcal{U}_{\text{R}}(\Psi)$, as defined by Eq. (A21), is the probability of getting the outcome $\hat{U}$ when the embedded unitary is U. $p_j(\hat{U}|U)$ is conjugate-invariant, and thus $p_j(\hat{U}|U) = p_j(U'^{\dagger}|I) = p_j(U'|I)$ with $U' := \hat{U}^{\dagger} U$. Substituting into the above expression, we have

$$\epsilon_j = \epsilon_{\text{wc}} \left(\int dU \int dU' \, p_j(U'|I) \mathcal{U}_{\text{L}} \circ \mathcal{U}_{\text{L}}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_{\text{C}}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1}, \mathcal{I}_{\text{L}} \right). \quad (A26)$$

Define the twirled version of the original protocol as

$$\mathcal{P}_{\text{twirl}} := \int dU \, \mathcal{U}_{\text{L}} \circ \mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1}. \quad (A27)$$

Protocol 1. Covariant implementation of $\mathcal{V}_{\text{L}}$ (see also Fig. 5).

Preparation: Initiate each reference register in the reference frame state Ψ.

Encoder $\mathcal{E}_{\text{cov}}$: Apply $\mathcal{U}_{\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1}$ to the logical system and $\mathcal{U}_{\text{R}}$ to the reference, with U following the Haar measure.

Gate implementation: Apply $\mathcal{V}_{\text{C}}$ on the computational system and $\mathcal{V}_{\text{R}}$ on the reference.

Error $\mathcal{C}$: An error may occur on the computational register C and the reference R.

Decoder $\mathcal{D}$: Measure the reference frame with the error-tolerant measurement $\mathcal{M}_{\hat{U}}$ from Eq. (A13). Depending on the measurement outcome $\hat{U}$, apply $\hat{\mathcal{U}}_{\text{L}} \circ \mathcal{D} \circ \hat{\mathcal{U}}_{\text{C}}^{-1}$ on the computational system.

Using lemma 5 in Appendix A 6, we bound the error as

$$\epsilon_j \leqslant 9d \cdot \max\{\tilde{\epsilon}_{j,\text{code}}, 1 - F_{j,\text{RF}}\}, \tag{A28}$$

where

$$F_{j,\text{RF}} := F_{\text{ent}}\left(\int dU \int dU' \, p_j(U'|I)\mathcal{U}_\text{L} \circ \mathcal{U}_\text{L}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_\text{C}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_\text{L}^{-1}, \mathcal{P}_{\text{twirl}}\right) \tag{A29}$$

is the entanglement fidelity for the reference frame correction and

$$\tilde{\epsilon}_{j,\text{code}} := \epsilon_{\text{wc}}(\mathcal{P}_{\text{twirl}}, \mathcal{I}_\text{L}) \tag{A30}$$

is the twirled code error. Notice that, by lemma 5, Eq. (A28) holds only if

$$\epsilon_{\text{ent}}\left(\int dU \int dU' \, p_j(U'|I)\mathcal{U}_\text{L} \circ \mathcal{U}_\text{L}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_\text{C}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_\text{L}^{-1}, \mathcal{P}_{\text{twirl}}\right) \leqslant \frac{1}{2}. \tag{A31}$$

By the Fuchs-van de Graaf inequalities Eq. (A6), this holds as long as $F_{j,\text{RF}} \geqslant 3/4$.

On one hand, by the joint convexity of the diamond norm error, the twirled code error is bounded by the code error, i.e.,

$$\tilde{\epsilon}_{j,\text{code}} \leqslant \epsilon_{j,\text{code}}. \tag{A32}$$

On the other hand, by the joint concavity of the square-root fidelity and the unitary invariance of fidelity, we have

$$\begin{aligned}
\sqrt{F}_{j,\text{RF}} &\geqslant \int dU \sqrt{F}_{\text{ent}}\left(\int dU' \, p_j(U'|I)\mathcal{U}_\text{L} \circ \mathcal{U}_\text{L}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_\text{C}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_\text{L}^{-1}, \mathcal{U}_\text{L} \circ \mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \circ \mathcal{U}_\text{L}^{-1}\right) \\
&= \int dU \sqrt{F}_{\text{ent}}\left(\int dU' \, p_j(U'|I)\mathcal{U}_\text{L}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_\text{C}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E}, \mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E}\right) \\
&= \sqrt{F}_{\text{ent}}\left(\int dU' \, p_j(U'|I)\mathcal{U}_\text{L}'^{-1} \circ \mathcal{D} \circ \mathcal{U}_\text{C}' \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E}, \mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E}\right).
\end{aligned} \tag{A33}$$

Using the maximally entangled state to "wire around" the channels, we get

$$\begin{aligned}
F_{j,\text{RF}} &\geqslant F\left(\int dU \, p_j(U|I)(\mathcal{D} \circ \mathcal{U}_\text{C} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \otimes \mathcal{U}_\text{L}^*)(\Phi_\text{L}^+), (\mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \otimes \mathcal{I}_\text{R})(\Phi_\text{L}^+)\right) \\
&\geqslant F_{\text{wc}}\left(\int dU \, p_j(U|I)(\mathcal{D} \circ \mathcal{U}_\text{C} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \otimes \mathcal{U}_\text{L}^*), \mathcal{D} \circ \mathcal{C}_{j,\text{C}} \circ \mathcal{E} \otimes \mathcal{I}_\text{L}\right).
\end{aligned} \tag{A34}$$

Here U^* denotes the complex conjugation of U. Exploiting the data processing inequality of fidelity, we have

$$F_{j,\text{RF}} \geqslant F_{\text{wc}}\left(\int dU \, p_j(U|I)\mathcal{U}_\text{C} \otimes \mathcal{U}_\text{L}^*, \mathcal{I}_\text{C} \otimes \mathcal{I}_\text{L}\right). \tag{A35}$$

Finally, combining Eqs. (A28), (A32), and (A35) gives (A19). ∎

The bound in lemma 1 applies to covariant codes constructed from arbitrary reference frames. For reference frames as in Eq. (A9), we can give a more detailed bound. Since $\epsilon_{j,\text{code}}$ is specified by the noncovariant code $(\mathcal{E}, \mathcal{D})$ that we use as a subroutine, what we need to do is to bound the reference frame error by determining

$$F_{\text{wc}}\left(\int dU \, p_j(U|I)\mathcal{U}_\text{C} \otimes \mathcal{U}_\text{L}^*, \mathcal{I}_\text{C} \otimes \mathcal{I}_\text{L}\right), \tag{A36}$$

where $p_j(U|I)$ is the distribution defined by Eq. (A21).

From now on we focus on erasure errors, which are indeed covariant and detectable. For erasure errors, as the reference frame register is initiated in multiple copies of the state ψ defined by Eq. (A9), the error map $\mathcal{C}_{j,\text{R}}$ will ruin some of the copies, and the remaining copies can still be cast in the form Eq. (A9). For simplicity, we still denote the distribution [as in

Eq. (A9)] associated to the state of the remaining copies by $\{q_\lambda\}$, keeping in mind that this distribution may depend on j. Then, by using Eq. (A12) we can write $p_j(U|I)$ as

$$p(U|I) = \left|\sum_{\lambda \in \text{S}_{\text{via}}} \sqrt{q_\lambda} \chi_{U^{-1},\lambda}\right|^2, \tag{A37}$$

where S_{via} is the "viable" set of Young diagrams on which $q_\lambda > 0$.

Define $n' := n_\text{C} + d - 1$ and the "interior" subset of S_{via} as

$$\text{S}_{\text{int}} := \{\lambda \in \text{S}_{\text{via}} : |\lambda_i - \lambda_j| > 3n' \, \forall \, i \neq j\}. \tag{A38}$$

Then we can show that

Lemma 2. For reference frames of the form in Eq. (A9),

$$\begin{aligned}
F_{\text{wc}}&\left(\int dU \, p_j(U|I)\mathcal{U}_\text{C} \otimes \mathcal{U}_\text{L}^*, \mathcal{I}_\text{C} \otimes \mathcal{I}_\text{L}\right) \\
&\geqslant \min_{\boldsymbol{\Delta} \in \text{S}_{\text{diff}}} \sum_{\lambda \in \text{S}_{\text{int}}} \sqrt{q_\lambda q_{\lambda+\boldsymbol{\Delta}}}, \tag{A39}
\end{aligned}$$

where $\text{S}_{\text{diff}} = \{\boldsymbol{\Delta} \in \mathbb{Z}^{\times d} : |\Delta_i| \leqslant n', i = 1, \ldots, d; \sum_{j=1}^d \Delta_j = 0\}$.

Proof. Denote by S_{cost} the collection of all Young diagrams appearing in the decomposition of $\mathcal{U}_{\text{L}}^* \otimes \mathcal{U}_{\text{C}}$, i.e., the set of Young diagrams corresponding to the "cost" function. Notice that, since the computational register consists of $n_{\text{C}} = n - n_{\text{R}}$ qudits and the logical register consists of one qudit, S_{cost} does not contain any $\boldsymbol{\lambda}$ with more than $n' := n_{\text{C}} + d - 1$ boxes. We have the decomposition

$$\mathcal{U}_{\text{L}}^* \otimes \mathcal{U}_{\text{C}} \simeq \bigoplus_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{cost}}} U_{\boldsymbol{\lambda}} \otimes I_{m_{\boldsymbol{\lambda}}} \tag{A40}$$

where $m_{\boldsymbol{\lambda}}$ are the multiplicities.

We now convert the quantity Eq. (A36) into a form that is easier to bound. For conjugate-invariant $p_j(U|I)$, the channel of interest $\int dU \, p_j(U|I) \bigoplus_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{cost}}} (\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathcal{I}_{m_{\boldsymbol{\lambda}}})$ is block-covariant with respect to the symmetry $\bigoplus_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{cost}}} (\mathcal{V}_{\boldsymbol{\lambda}} \otimes \mathcal{I}_{m_{\boldsymbol{\lambda}}})$ for $V \in \mathsf{SU}(d)$. We can now apply the following lemma (see Sec. A 7 for the proof):

Lemma 3. The worst-case input fidelity of a channel $\mathcal{E}$ commuting with unitary channels of the block diagonal form $\bigoplus_{\boldsymbol{\lambda}} (\mathcal{U}_{\boldsymbol{\lambda}} \otimes \mathcal{I}_{m_{\boldsymbol{\lambda}}})$ for any $U \in \mathsf{SU}(d)$ can be achieved by an input state of the following form:

$$|\Psi^*\rangle := \bigoplus_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}} |\Phi_{\boldsymbol{\lambda}}^+\rangle \otimes |\psi_{m_{\boldsymbol{\lambda}}}\rangle \tag{A41}$$

with $|\Phi_{\boldsymbol{\lambda}}^+\rangle$ being the maximally entangled state in $\mathcal{H}_{\boldsymbol{\lambda}} \otimes \mathcal{H}_{\boldsymbol{\lambda}}'$, $\{c_{\boldsymbol{\lambda}}\}$ being an amplitude distribution, and $|\psi_{m_{\boldsymbol{\lambda}}}\rangle$ being a fixed (otherwise arbitrary) state on the multiplicity subspace.

Then, the worst-case input fidelity Eq. (A36) can be achieved by a state of the form Eq. (A41).

Notice that, since the error $\mathcal{C}_{j,\text{R}}$ simply destroys a few copies of the reference state, the remaining copies can still be cast in the form Eq. (A10). Therefore $p_j(U|I)$ is of the form Eq. (A12). We now combine the above facts with lemma 3 and

express the fidelity Eq. (A36) as

$$F_{\text{wc}}\left(\int dU \, p_j(U|I) \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{L}}^*, \mathcal{I}_{\text{C}} \otimes \mathcal{I}_{\text{L}}\right)$$
$$= \sum_{\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \mathsf{S}_{\text{via}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}'}} S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'}, \tag{A42}$$

where S_{via} is the set of Young diagrams on which $q_{\boldsymbol{\lambda}} > 0$, $S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'}$ is a correlation function defined as

$$S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'} := \int dU \, \chi_{U, \boldsymbol{\lambda}} \chi_{U, \boldsymbol{\lambda}'}^* \left| \sum_{\boldsymbol{\mu} \in \mathsf{S}_{\text{cost}}} |c_{\boldsymbol{\mu}}|^2 d_{\boldsymbol{\mu}}^{-1} \chi_{U, \boldsymbol{\mu}} \right|^2 \tag{A43}$$

and $\{c_{\boldsymbol{\mu}}\}$ is an amplitude distribution over S_{cost}. Notice that here $\{q_{\boldsymbol{\lambda}}\}$ corresponds to the reference frame state when the error syndrome $\mathcal{C}_{j,\text{R}}$ takes place, and we abbreviated the index j for simplicity.

By the orthogonality of the characters, $S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'} \geqslant 0$ depends on the overlap between the irreducible decomposition of $\boldsymbol{\lambda} \otimes \boldsymbol{\mu}$ and the irreducible decomposition of $\boldsymbol{\lambda}' \otimes \boldsymbol{\mu}'$. Since S_{cost} contains only diagrams with no more than n' boxes, $S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'} = 0$ unless

$$d_{\text{Young}}(\boldsymbol{\lambda}, \boldsymbol{\lambda}') \leqslant n', \tag{A44}$$

where $d_{\text{Young}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \frac{1}{2} \sum_{i=1}^{d} |\lambda_i - \mu_i|$ is the distance between Young diagrams.

Now we turn back to the fidelity. First, since $S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'} \geqslant 0$ with equality for $d_{\text{Young}}(\boldsymbol{\lambda}, \boldsymbol{\lambda}') > n'$, we have

$$F_{\text{wc}} \geqslant \sum_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}} \sum_{\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}+\boldsymbol{\Delta}}} S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}}, \tag{A45}$$

where S_{int} is the interior subset of S_{via} defined by Eq. (A38). Crucially, we now argue that for $\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}$ and $\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}$, the correlation function depends only on their relative distance $\boldsymbol{\Delta}$, but not explicitly on $\boldsymbol{\lambda}$, i.e.,

$$S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}} = \tilde{S}_{\boldsymbol{\Delta}}. \tag{A46}$$

Invoking Eq. (A43), the correlation function can be expressed as

$$S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}} = \sum_{\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathsf{S}_{\text{cost}}} |c_{\boldsymbol{\mu}} c_{\boldsymbol{\mu}'}|^2 (d_{\boldsymbol{\mu}} d_{\boldsymbol{\mu}'})^{-1} \int dU \, \chi_{U, \boldsymbol{\lambda}} \chi_{U, \boldsymbol{\mu}} \chi_{U, \boldsymbol{\lambda}+\boldsymbol{\Delta}}^* \chi_{U, \boldsymbol{\mu}'}^*. \tag{A47}$$

By orthogonality of the characters, the correlation function is just

$$S_{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}} = \sum_{\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathsf{S}_{\text{cost}}} |c_{\boldsymbol{\mu}} c_{\boldsymbol{\mu}'}|^2 (d_{\boldsymbol{\mu}} d_{\boldsymbol{\mu}'})^{-1} C_{\boldsymbol{\mu}, \boldsymbol{\mu}'}^{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}}, \tag{A48}$$

$$C_{\boldsymbol{\mu}, \boldsymbol{\mu}'}^{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}} = \left| \left\{ \left(L_{\boldsymbol{\nu}/\boldsymbol{\lambda}}^{\boldsymbol{\mu} \to \tilde{\boldsymbol{\mu}}}, L_{\boldsymbol{\nu}/(\boldsymbol{\lambda}+\boldsymbol{\Delta})}^{\boldsymbol{\mu}' \to \tilde{\boldsymbol{\mu}}'} \right) : \boldsymbol{\nu} \in \boldsymbol{\lambda} \otimes \boldsymbol{\mu}, \, \boldsymbol{\nu} \in (\boldsymbol{\lambda} + \boldsymbol{\Delta}) \otimes \boldsymbol{\mu}' \right\} \right|. \tag{A49}$$

Here $L_{\boldsymbol{\nu}/\boldsymbol{\lambda}}^{\boldsymbol{\mu} \to \tilde{\boldsymbol{\mu}}}$ denotes a Littlewood-Richardson tableau of shape $\boldsymbol{\nu}/\boldsymbol{\lambda}$ with content $\tilde{\boldsymbol{\mu}}$, obtained by adding the Young diagram $\boldsymbol{\mu}$ to $\boldsymbol{\lambda}$ according to the Littlewood-Richardson rule, and $\boldsymbol{\nu} \in \boldsymbol{\lambda} \otimes \boldsymbol{\mu}$ means that $\boldsymbol{\nu}$ appears at least once in the decomposition of $\boldsymbol{\lambda} \otimes \boldsymbol{\mu}$. Now one can see why Eq. (A46) holds. Indeed, S_{int} is so defined [cf. Eq. (A38)] that, when $\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}$, the lengths of different rows of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda} + \boldsymbol{\Delta}$ have big enough gaps so that even adding all n' boxes to one row would not make its box number greater than its preceding rows. Therefore, according to the Littlewood-Richardson rule adding $\boldsymbol{\mu}'$ and $\boldsymbol{\mu}$, neither has more than n' boxes, is not constraint by the shape of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda} + \boldsymbol{\Delta}$. The sum $\sum_{\boldsymbol{\Delta}} C_{\boldsymbol{\mu}, \boldsymbol{\mu}'}^{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}}$ is determined by how many different contents can any $\boldsymbol{\mu} \in \mathsf{S}_{\text{cost}}$ possibly generate.

The following property will be useful: rectifications of Littlewood-Richardson tableaux and representative Young tableaux (of symmetry tensors) are in one-to-one correspondence. Indeed, every representative Young tableau is a standard (i.e., left $\leqslant$ right and top $<$ bottom) tableau. It cor-

responds to a rectified Littlewood-Richardson tableau, whose jth row has a number x of the index i with x being the number of js in the ith row of the representative Young tableau. A rectified Littlewood-Richardson tableau of content $\tilde{\boldsymbol{\mu}}$ thus corresponds to the representative Young tableau $\tilde{\boldsymbol{\mu}}$. Since the total number of representative Young tableaux is the dimension of the irreducible representation, the total number of contents that $\boldsymbol{\mu} \in \mathsf{S}_{\text{cost}}$ can generate is $d_{\boldsymbol{\mu}}$. Therefore we have

$$\sum_{\boldsymbol{\Delta}} C_{\boldsymbol{\mu}, \boldsymbol{\mu}'}^{\boldsymbol{\lambda}, \boldsymbol{\lambda}+\boldsymbol{\Delta}} = d_{\boldsymbol{\mu}} d_{\boldsymbol{\mu}'} \tag{A50}$$

for any $\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathsf{S}_{\text{cost}}$. Combining with Eqs. (A46) and (A48), we have

$$\sum_{\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}} \tilde{S}_{\boldsymbol{\Delta}} = \sum_{\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathsf{S}_{\text{cost}}} |c_{\boldsymbol{\mu}} c_{\boldsymbol{\mu}'}|^2 = 1. \tag{A51}$$

Now, the fidelity bound Eq. (A45) becomes

$$F_{\text{wc}}\left(\int dU \, p_j(U|I) \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{L}}^*, \mathcal{I}_{\text{C}} \otimes \mathcal{I}_{\text{L}} \right)$$
$$\geqslant \sum_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}} \sum_{\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}+\boldsymbol{\Delta}}} \tilde{S}_{\boldsymbol{\Delta}}$$
$$\geqslant \left(\min_{\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}} \sum_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}+\boldsymbol{\Delta}}} \right) \left(\sum_{\boldsymbol{\Delta}' \in \mathsf{S}_{\text{diff}}} \tilde{S}_{\boldsymbol{\Delta}'} \right)$$
$$= \min_{\boldsymbol{\Delta} \in \mathsf{S}_{\text{diff}}} \sum_{\boldsymbol{\lambda} \in \mathsf{S}_{\text{int}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}+\boldsymbol{\Delta}}}, \tag{A52}$$

and the proof is complete. ∎

From the above discussion, we obtain a bound (by combining lemma 1 with lemma 2) on the performance of our protocol. In the following sections, we will use it to evaluate the performance of our protocol for two different erasure error models.

(1) *Weak erasure error* (Sec. B). At most n_{e} out of the n qudits are erased, with n_{e} a constant independent of n.

(2) *i.i.d. erasure error* (Sec. C). Each qudit has a constant probability of being erased. The errors on different qudits are independent.

5. A lower bound on ϵ_{cov}

In the last part of this section, we derive a lower bound on ϵ_{cov} of any covariant code, which will be used to show the optimality of our protocol. The lower bound is built upon the main result of Ref. [47], further strengthened here by us. This bound was derived for $\mathsf{U}(1)$, but we can always find a single-parameter family of unitaries embedded in $\mathsf{SU}(d)$, so the bound applies in general. For instance, we can consider the $\mathsf{U}(1)$ family $U_\theta := e^{-i\theta H}$ generated by the Hamiltonian

$$H = \sum_{j=1}^{d} h_j |j\rangle\langle j| \tag{A53}$$

for an orthonormal basis $\{|j\rangle\}_{j=1}^{d}$ of $\mathcal{H}$. The $\mathsf{SU}(d)$ covariance implies the $\mathsf{U}(1)$ covariance, i.e., $U_\theta^{\otimes n} \circ \mathcal{E}_{\text{cov}} = \mathcal{E}_{\text{cov}} \circ \mathcal{U}_\theta$ for any θ.

In Ref. [47], it was shown for covariant codes the worst-case input fidelity obeys the lower bound:

$$\sqrt{1 - F_{\text{wc}}} \geqslant \frac{(\Delta H)^2}{3\sqrt{6} I_{\text{Fisher}}^\uparrow}. \tag{A54}$$

Here ΔH is the difference between the maximum eigenvalue and the minimum eigenvalue of H, and $I_{\text{Fisher}}^\uparrow$ is a Fisher information upper bound of the channel $\mathcal{C}_\theta := \mathcal{C} \circ \mathcal{U}_\theta^{\otimes n}$ (with $\mathcal{C}$ being the error) that can be evaluated as follows: if there exists Kraus operators $\{K_{l;\theta}\}$ of the channel $\mathcal{C}_\theta$ such that $\sum_l \dot{K}_{l;\theta}^\dagger K_{l;\theta} = 0$ (here $\dot{K}_{l;\theta}$ denotes the derivative of $K_{l;\theta}$ with respect to θ), the Fisher information upper bound is again bounded as

$$I_{\text{Fisher}}^\uparrow \leqslant 4 \left\| \sum_l \dot{K}_{l;\theta}^\dagger \dot{K}_{l;\theta} \right\|_\infty. \tag{A55}$$

Here $\| \cdot \|_\infty$ denotes the operator norm. If there does not exist such a Kraus form, we set $I_{\text{Fisher}}^\uparrow = \infty$ and the bound Eq. (A54) is trivial.

The above bound, however, is not enough to show the optimality of our result. Instead, we derive a strengthened version of it.

Lemma 4 (Theorem 1 of Ref. [47]; strengthened version). The error of any covariant code is lower bounded by

$$\epsilon_{\text{cov}} \geqslant \frac{(\Delta H)^2}{16 I_{\text{Fisher}}^\uparrow}. \tag{A56}$$

Here ΔH is the difference between the maximum eigenvalue and the minimum eigenvalue of H, and $I_{\text{Fisher}}^\uparrow$ is the Fisher information upper bound Eq. (A55).

To see Eq. (A56) is indeed a strengthening of Eq. (A54), simply notice that $\epsilon_{\text{cov}} \leqslant \sqrt{1 - F_{\text{wc}}}$ [80] and thus Eq. (A56) implies Eq. (A54). On the other hand, plugging the other part of the Fuchs-van de Graaf inequality $\epsilon_{\text{cov}} \geqslant 1 - \sqrt{F_{\text{wc}}}$ into Eq. (A54) only yields a bound on ϵ_{cov} that scales as $(I_{\text{Fisher}}^\uparrow)^{-2}$ which, as we will soon see, is not enough to prove the desired $1/n^2$-scaling bound on ϵ_{cov}. For this reason, we must use the strengthened version.

Proof of lemma 4. This result can be derived along the same line of arguments as the proof of the original bound Eq. (A54) in Ref. [47]. A few improvements need to be made as the following: First, in Eq. (A11) of Ref. [47], we directly consider the worst-case error ϵ_{cov} and employ the tighter bound (than the Fuchs-van de Graaf inequality) between f^2 an ϵ_{wc} when one of the two states is pure; see, e.g., ([84], Eq. (9.111)). Then Eq. (A11) becomes

$$f^2 \geqslant 1 - \epsilon_{\text{cov}}, \tag{A57}$$

where f^2 is the same quantity as in the original equation. With this improved inequality substituted into Eq. (A20), we get the counterpart of ([47], Lemma 1) for ϵ_{wc}:

$$I_{\text{Fisher}}^\uparrow \geqslant (1 - 4\epsilon_{\text{cov}})(\Delta H)^2. \tag{A58}$$

Equation (21) of [47] then becomes $(m\Delta H)^2(1 - 4m\epsilon_{\text{cov}}) \leqslant m I_{\text{Fisher}}^\uparrow$ and optimizing over m we get the bound (A56). ∎

6. On the error of covariant channels

Let $\mathcal{A}$ and $\mathcal{B}$ be channels acting on a d-dimensional Hilbert space $\mathcal{H}$. Assume that $\mathcal{A}$ and $\mathcal{B}$ are both covariant with the (full) symmetry group $\mathsf{SU}(d)$ on $\mathcal{H}$. Define the entanglement error:

$$\epsilon_{\text{ent}}(\mathcal{X}, \mathcal{Y}) := \tfrac{1}{2}\|X - Y\|_1,$$

where X and Y are the Choi states of two channels $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Here we prove the following lemma:

Lemma 5. Suppose $\epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B}) \leqslant 1/2$. The worst-case input error of $\mathcal{B}$ can be bounded as

$$\epsilon_{\text{wc}}(\mathcal{B}, \mathcal{I}) \leqslant 9d \cdot \max\{\epsilon_{\text{ent}}(\mathcal{A}, \mathcal{I}), 1 - F_{\text{ent}}(\mathcal{A}, \mathcal{B})\}. \quad \text{(A59)}$$

Here $\mathcal{I}$ denotes the identity channel on $\mathcal{H}$.

Proof. In the following, we make frequent use of an elementary relation between the worst-case input error and the entanglement error:

$$\epsilon_{\text{wc}}(\mathcal{X}, \mathcal{Y}) \leqslant d \cdot \epsilon_{\text{ent}}(\mathcal{X}, \mathcal{Y}). \quad \text{(A60)}$$

For any channel $\mathcal{A}$, define its Choi state as

$$A := (\mathcal{A} \otimes \mathcal{I})(\Phi_d^+) \quad \text{(A61)}$$

with Φ_d^+ being the maximally entangled state in $\mathcal{H} \otimes \mathcal{H}$. When $\mathcal{A}$ is covariant, we have

$$[A, U \otimes U^*] = 0, \quad \forall U \in \mathsf{SU}(d). \quad \text{(A62)}$$

By Schur's lemma, the Choi states of the covariant channels $\mathcal{A}$ and $\mathcal{B}$ can be decomposed as

$$A = (1 - a) \cdot \Phi_d^+ + a \cdot \rho^\perp \quad \rho^\perp := \frac{1}{d^2 - 1}(I \otimes I - \Phi_d^+),$$
$$\text{(A63)}$$

$$B = (1 - b) \cdot \Phi_d^+ + b \cdot \rho^\perp. \quad \text{(A64)}$$

Therefore the entanglement error and the entanglement fidelity that we are interested in can be evaluated as

$$\epsilon_{\text{ent}}(\mathcal{A}, \mathcal{I}) = a, \quad \text{(A65)}$$

$$\epsilon_{\text{ent}}(\mathcal{B}, \mathcal{I}) = b, \quad \text{(A66)}$$

$$F_{\text{ent}}(\mathcal{A}, \mathcal{B}) = (\sqrt{(1 - a)(1 - b)} + \sqrt{ab})^2, \quad \text{(A67)}$$

$$\epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B}) = |a - b|. \quad \text{(A68)}$$

If $a \geqslant b$, we have

$$\epsilon_{\text{wc}}(\mathcal{B}, \mathcal{I}) \leqslant d \cdot \epsilon_{\text{ent}}(\mathcal{B}, \mathcal{I}) = d \cdot b \leqslant d \cdot a. \quad \text{(A69)}$$

If $a < b$, we can write $b = a + \epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B})$. We further distinguish between two cases: If $a > \epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B})/8$, then $b \leqslant 9a$ and thus

$$\epsilon_{\text{wc}}(\mathcal{B}, \mathcal{I}) \leqslant 9d \cdot a. \quad \text{(A70)}$$

Otherwise, if $a \leqslant \epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B})/8$, we have (using the shorthand $\epsilon_{\text{ent}} := \epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B})$)

$$1 - F_{\text{ent}}(\mathcal{A}, \mathcal{B})$$
$$= 1 - (\sqrt{(1 - a)(1 - a - \epsilon_{\text{ent}})} + \sqrt{a(a + \epsilon_{\text{ent}})})^2$$
$$= \epsilon_{\text{ent}} + 2a(1 - a - \epsilon_{\text{ent}})$$

$$- 2\sqrt{a(1 - a)(a(1 - a) + \epsilon_{\text{ent}}(1 - 2a - \epsilon_{\text{ent}}))}$$
$$\geqslant \epsilon_{\text{ent}} + 2a(1 - a - \epsilon_{\text{ent}}) - 2\sqrt{2a(1 - a)\epsilon_{\text{ent}}}$$
$$= (\sqrt{\epsilon_{\text{ent}}} - \sqrt{2a(1 - a)})^2 - 2a\epsilon_{\text{ent}}$$
$$\geqslant \left(\sqrt{\epsilon_{\text{ent}}} - \frac{\sqrt{\epsilon_{\text{ent}}}}{2}\right)^2 - \frac{\epsilon_{\text{ent}}^2}{4}$$
$$\geqslant \frac{\epsilon_{\text{ent}}(1 - \epsilon_{\text{ent}})}{4} \geqslant \frac{\epsilon_{\text{ent}}}{8}.$$

Notice that the last inequality holds if $\epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B}) \leqslant 1/2$. Then, the worst-case input error of $\mathcal{B}$ can be bounded as

$$\epsilon_{\text{wc}}(\mathcal{B}, \mathcal{I}) \leqslant d \cdot (a + \epsilon_{\text{ent}}(\mathcal{A}, \mathcal{B}))$$
$$\leqslant d \cdot (a + 8(1 - F_{\text{ent}}(\mathcal{A}, \mathcal{B}))). \quad \text{(A71)}$$

Finally, we get the desired bound by summarizing Eqs. (A69), (A70), and (A71) into a more compact form. ∎

7. Proof of lemma 3

Here we show that the worst-case input fidelity of a channel $\mathcal{E}$ commuting with

$$\mathcal{U}_{\text{tot}} := \bigoplus_\lambda \mathcal{U}_\lambda \otimes \mathcal{I}_{m_\lambda} \quad \text{(A72)}$$

can be achieved by an input state of the following form:

$$|\Psi^*\rangle := \bigoplus_\lambda c_\lambda |\Phi_\lambda^+\rangle \otimes |\psi_{m_\lambda}\rangle. \quad \text{(A73)}$$

The worst-case input fidelity of $\mathcal{E}$ can be written as a SDP (with the Slater's condition always satisfied). In particular, from lemma 6 (see later) the primal problem for the worst-case input (square-root) fidelity is

$$\sqrt{F}_{\text{wc}}(\mathcal{E}, \mathcal{I}) = \min \tfrac{1}{2}(\text{Tr}(E_{BA}\Gamma_{BA}) + \text{Tr}(\Omega_{BA}\Lambda_{BA}))$$

$$\text{such that } \begin{pmatrix} \Gamma_{BA} & -I_B \otimes \rho_A^T \\ -I_B \otimes \rho_A^T & \Lambda_{BA} \end{pmatrix} \geqslant 0$$

$$\text{Tr}(\rho_A^T) \geqslant 1$$

$$\rho_A, \Gamma_{BA}, \Lambda_{BA} \geqslant 0. \quad \text{(A74)}$$

Here $\rho_A := \text{Tr}_R \Psi_{AR}$ corresponds to the marginal of the input state, $\Omega_{BA} := |I\rangle\!\rangle\langle\!\langle I|$, and $E_{BA} := (\mathcal{E} \otimes \mathcal{I}_A)(\Omega)$.

Suppose that $|\Psi\rangle$ with marginal ρ_A is an input state achieving the minimum. Consider the twirling

$$\mathcal{T}(\cdot) := \int dU \left(\bigoplus_\lambda \mathcal{U}_\lambda \otimes \mathcal{U}_\lambda^* \otimes \mathcal{I}_{m_\lambda} \otimes \mathcal{I}_{m_\lambda}\right)(\cdot) \quad \text{(A75)}$$

on the constraints. The constraints become

$$\begin{pmatrix} \mathcal{T}(\Gamma_{BA}) & -I_B \otimes \tilde{\rho}_A^T \\ -I_B \otimes \tilde{\rho}_A^T & \mathcal{T}(\Lambda_{BA}) \end{pmatrix} \geqslant 0, \quad \text{(A76)}$$

$$\text{Tr}(\tilde{\rho}_A^T) \geqslant 1, \quad \text{(A77)}$$

$$\tilde{\rho}_A, \mathcal{T}(\Gamma_{BA}), \mathcal{T}(\Lambda_{BA}) \geqslant 0. \quad \text{(A78)}$$

Here $\tilde{\rho}_A := \int dU \bigoplus_\lambda \mathcal{U}_\lambda^* \otimes \mathcal{I}_{m_\lambda}(\rho_A)$. Noticing that both E_{BA} and Ω_{BA} are invariant under $\mathcal{T}$ and its dual $\mathcal{T}^\dagger$, the objective

function remains the same:

$$\tfrac{1}{2}(\mathrm{Tr}(E_{BA}\mathcal{T}(\Gamma_{BA})) + \mathrm{Tr}(\Omega_{BA}\mathcal{T}(\Lambda_{BA})))$$
$$= \tfrac{1}{2}(\mathrm{Tr}(E_{BA}\Gamma_{BA}) + \mathrm{Tr}(\Omega_{BA}\Lambda_{BA})). \qquad (A79)$$

Therefore an input $|\Psi^*\rangle$ with marginal $\tilde{\rho}_A$ also achieves the worst-case input fidelity. Since it is invariant under $\bigoplus_\lambda \mathcal{U}_\lambda^*$, by Schur's lemma it has to be of the diagonal form

$$\tilde{\rho}_A = \bigoplus_\lambda |c_\lambda|^2 I_\lambda \otimes \sigma_{m_\lambda} \qquad (A80)$$

with $\{\sigma_{m_\lambda}\}$ being arbitrary states on the multiplicity subspaces. Therefore the corresponding input state $|\Psi^*\rangle$ can be cast into the desired form Eq. (A73).

At last we prove the formulation Eq. (A74) of the worst-case input fidelity as a semidefinite program, adapting the method used for the diamond norm [85]. First, for a channel $\mathcal{A}: \mathsf{Lin}(\mathcal{K}_\mathrm{in}) \to \mathsf{Lin}(\mathcal{K}_\mathrm{out})$, define its Choi operator $A \in \mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}'_\mathrm{in})$ with $\mathcal{K}'_\mathrm{in} \simeq \mathcal{K}_\mathrm{in}$ as follows. For $|\Omega\rangle \in \mathcal{K}_\mathrm{in} \otimes \mathcal{K}'_\mathrm{in}$ the unnormalized maximally entangled state, set $A = (\mathcal{A} \otimes \mathcal{I})(\Omega)$.

Lemma 6. Given any two channels $\mathcal{A}$ and $\mathcal{B}$ with Choi operators A and B, respectively,

$$\sqrt{F_\mathrm{wc}(\mathcal{A}, \mathcal{B})} = \min \ \tfrac{1}{2}(\mathrm{Tr}[A\Gamma] + \mathrm{Tr}[B\Lambda])$$

$$\text{s.t.} \ \begin{pmatrix} \Gamma & -I \otimes \rho \\ -I \otimes \rho & \Lambda \end{pmatrix} \geqslant 0,$$

$$\rho, \Gamma, \Lambda \geqslant 0,$$

$$\rho \in \mathsf{St}(\mathcal{K}_\mathrm{in}),$$

$$\Gamma, \Lambda \in \mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}_\mathrm{in}). \qquad (A81)$$

Proof. The first step in the proof is to simplify the dependence on the input state. Using the dual form of the fidelity function from Refs. [85,86] in (A1), we have

$$\sqrt{F_\mathrm{wc}(\mathcal{A}, \mathcal{B})}$$
$$= \min \ \tfrac{1}{2}(\mathrm{Tr}[Y(\mathcal{A} \otimes \mathcal{I}_\mathrm{R})(\rho)] + \mathrm{Tr}[Z(\mathcal{B} \otimes \mathcal{I}_\mathrm{R})(\rho)])$$

$$\text{s.t.} \ \begin{pmatrix} Y & -I \\ -I & Z \end{pmatrix} \geqslant 0,$$
$$\rho \in \mathsf{St}(\mathcal{K}_\mathrm{in} \otimes \mathcal{K}_\mathrm{R}),$$
$$Y, Z \in \mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}_\mathrm{R}). \qquad (A82)$$

Observe that we can assume the optimal input state ρ is pure because the objective function is linear in ρ. Therefore we can write it as $\rho = K\Omega K^\dagger$, where $\Omega \in \mathsf{Lin}(\mathcal{K}_\mathrm{in} \otimes \mathcal{K}'_\mathrm{in})$ and $K: \mathcal{K}'_\mathrm{in} \to \mathcal{K}_\mathrm{R}$ is the operator defined by $|\psi\rangle \mapsto \langle\Omega|(|\rho\rangle \otimes |\psi\rangle)$. Now let K be the variable in the optimization. By construction, $\rho \geqslant 0$, and the trace constraint becomes $\mathrm{Tr}[K^\dagger K] = 1$. Note that $K^\dagger K \in \mathsf{Lin}(\mathcal{H}'_\mathrm{in})$.

In the first term of the objective function, we recognize the Choi operator of $\mathcal{A}$:

$$\mathrm{Tr}[Y(\mathcal{A} \otimes \mathcal{I}_\mathrm{R})(\rho)] = \mathrm{Tr}[YK(\mathcal{A} \otimes \mathcal{I})(\Omega)K^\dagger] \qquad (A83)$$

$$= \mathrm{Tr}[K^\dagger YKA], \qquad (A84)$$

and this works similarly for the second term. Thus the optimization takes the form

$$\sqrt{F_\mathrm{wc}(\mathcal{A}, \mathcal{B})} = \min \ \tfrac{1}{2}(\mathrm{Tr}[K^\dagger YKA] + \mathrm{Tr}[K^\dagger ZKB])$$

$$\text{s.t.} \ \begin{pmatrix} Y & -I \\ -I & Z \end{pmatrix} \geqslant 0,$$
$$\mathrm{Tr}[K^\dagger K] = 1,$$
$$K \in \mathsf{Lin}(\mathcal{K}'_\mathrm{in} \to \mathcal{K}_\mathrm{R}),$$
$$Y, Z \in \mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}_\mathrm{R}). \qquad (A85)$$

Now we want to move to different variables, but without changing the optimal value. Defining Γ and Λ in $\mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}'_\mathrm{in})$ by $\Gamma = K^\dagger YK$ and $\Lambda = K^\dagger ZK$, the objective function becomes $\tfrac{1}{2}(\mathrm{Tr}[\Gamma A] + \mathrm{Tr}[\Lambda B])$. To deal with the constraints, let $M = \mathrm{diag}(K, K)$ and conjugate the block matrix in the constraint by M, multiplying from the left by $M^\dagger$ and the right by M. Doing so puts Γ and Λ on the diagonal, and $-I \otimes K^\dagger K$ on the off-diagonal. Note that the block matrix constraint implies $Y \geqslant 0$ and $Z \geqslant 0$, which by construction then implies $\Gamma \geqslant 0$ and $\Lambda \geqslant 0$. Defining $\rho \in \mathsf{Lin}(\mathcal{K}'_\in)$ as $\rho = K^\dagger K$, we have an optimization in the variables Γ, Λ and ρ. However, conjugation generally relaxes the constraints, which could lead to a smaller minimum value than the original optimization. Ignoring the difference between $\mathcal{K}'_\mathrm{in}$ and $\mathcal{K}_\mathrm{in}$, we have established

$$\sqrt{F(\mathcal{A}, \mathcal{B})} \leqslant \min \ \tfrac{1}{2}(\mathrm{Tr}[\Gamma A] + \mathrm{Tr}[\Lambda B])$$

$$\text{s.t.} \ \begin{pmatrix} \Gamma & -I \otimes \rho \\ -I \otimes \rho & \Lambda \end{pmatrix} \geqslant 0,$$

$$\rho, \Gamma, \Lambda \geqslant 0,$$

$$\rho \in \mathsf{St}(\mathcal{K}_\mathrm{in}),$$

$$\Gamma, \Lambda \in \mathsf{Lin}(\mathcal{K}_\mathrm{out} \otimes \mathcal{K}_\mathrm{in}). \qquad (A86)$$

To establish equality in Eq. (A86) and complete the proof, we show that any feasible variables in Eq. (A86) can be converted into feasible variables in Eq. (A85) having the same value of the objective function. Nominally, the following choice will work. Pick an arbitrary isometry $V: \mathcal{K}_\mathrm{in} \to \mathcal{K}_\mathrm{R}$ and define

$$K = V\rho^{1/2}, \qquad (A87)$$

$$Y = V\rho^{-1/2}\Gamma\rho^{-1/2}V^\dagger, \qquad (A88)$$

$$Z = V\rho^{-1/2}\Lambda\rho^{-1/2}V^\dagger. \qquad (A89)$$

However, the inverse of ρ is potentially problematic, as we do not know that ρ is full rank. We can avoid this problem as follows (which could presumably also be done by continuity). Suppose that $P \in \mathsf{Lin}(\mathcal{K}_\mathrm{in})$ is the projection onto the support of ρ. Then $\Gamma' = P\Gamma P$ and $\Lambda' = P\Lambda P$ are also feasible in Eq. (A86), as we can conjugate the constraints by P. Moreover, these variables will not have a larger value of the objective function, so we may as well begin the argument with feasible variables of this form.

Now we make a slight modification of the above choice, using the inverse on the support of ρ:

$$K = V\rho^{1/2}, \tag{A90}$$

$$Y = V(\rho^{-1/2}\Gamma\rho^{-1/2} + I \otimes (I - P))V^\dagger, \tag{A91}$$

$$Z = V(\rho^{-1/2}\Lambda\rho^{-1/2} + I \otimes (I - P))V^\dagger. \tag{A92}$$

Given that the support of Γ is contained in that of $I \otimes \rho$, it follows that $\mathrm{Tr}[K^\dagger Y K A] = \mathrm{Tr}[\Gamma A]$, and similarly for the other term. Hence this choice of variables leads to the same value of the objective function. Feasibility in Eq. (A85) also holds. The positivity and trace conditions hold immediately, and only the block matrix constraint is a little more involved. Define $P' \in \mathrm{Lin}(\mathcal{K}_R)$ by $P' = VPV^\dagger$. Then conjugating the block matrix in Eq. (A86) by $M = \mathrm{diag}(L, L)$ for $L = I \otimes V\rho^{-1/2}$ gives

$$\begin{pmatrix} V\rho^{-1/2}\Gamma\rho^{-1/2}V^\dagger & -I \otimes P' \\ -I \otimes P' & V\rho^{-1/2}\Lambda\rho^{-1/2}V^\dagger \end{pmatrix} \geqslant 0. \tag{A93}$$

To this inequality we can add

$$\begin{pmatrix} I \otimes (I - P') & -I \otimes (I - P') \\ -I \otimes (I - P') & I \otimes (I - P') \end{pmatrix} \geqslant 0, \tag{A94}$$

and the result is the block matrix constraint in Eq. (A85). ∎

Note that a different SDP for $F_{wc}(\mathcal{A}, \mathcal{I})$ (not the square root) was given in ([87], Eq. (A12)). After completion of this work, we discovered that the SDP of Lemma 6 also appears as Proposition 50 of [88].

APPENDIX B: HEISENBERG-LIMITED ERROR CORRECTION IN THE WEAK ERROR MODEL

1. Setting

In this section, we consider a relatively weak type of errors (compared to the other type we will consider). Specifically, the error is that at most n_e qudits among the n qudit systems composing the computational register and the reference frame register, are randomly lost:

$$\mathcal{C} = p_0\mathcal{I} + \sum_{\mathbf{s} \subset [n]:|\mathbf{s}| \leqslant n_e} p_\mathbf{s}(\mathcal{C}_e)_\mathbf{s}, \tag{B1}$$

where $(\mathcal{C}_e)_\mathbf{s}$ denotes the erasure of qudits whose labels are in the set $\mathbf{s}$ and $\{p_\mathbf{s}\}$ is a probability distribution.

Any quantum error-correcting code over qudits of distance at least $k + 1$ can perfectly correct k erasures [3]. For instance, the polynomial codes of Aharonov and Ben-Or are $[[2k+1, 1, k+1]]_d$ (one logical qudit encoded into $2k + 1$ computational qudits; code distance $k + 1$) stabilizer codes with this property [45]. In these cases, we can employ one of the perfect codes as the noncovariant subroutine $(\mathcal{E}, \mathcal{D})$ of our code. Since we only need to encode one logical qudit, the perfect code requires only $n_C = O(1)$ computational qudits.

This is to say: the (noncovariant) code $(\mathcal{E}, \mathcal{D})$ we require only $O(1)$ computational qudits and satisfies

$$\epsilon_{\mathbf{s}_C,\text{code}} = 0 \quad \forall |\mathbf{s}_C| \leqslant n_e, \tag{B2}$$

where $\mathbf{s}_C \subset [n_C]$ is the set of indices for error locations on the computational register. Now we arrange the reference frame register against this type of noise. To this purpose, we divide

the n_R qudit there into $n_e + 1$ groups, each consisting of $2m$ qudits, and we have the relation:

$$n_R = 2m(n_e + 1). \tag{B3}$$

On each group we construct a (highly coupled) reference frame state ψ of the form Eq. (A9). Then, at least one of the reference frame states will survive the erasure, and we can measure it to obtain the embedded rotation. In the following, we evaluate this (worst) case. Now we fix the form of the reference frame state by specifying the distribution $\{q_\lambda\}$ in Eq. (A9). We first specify $\mathsf{S}_{\text{via}} \subset \mathsf{Y}_m$ on which $q_\lambda > 0$. To this purpose, we first define a parameter M that depends on m as

$$M = \left\lfloor \frac{1}{3}\left(\frac{2m}{d(d-1)} - 1\right) \right\rfloor \tag{B4}$$

and $m_0 := m - d(d-1)(3M+1)/2$. Define $\tilde{\boldsymbol{\mu}} \in \mathsf{Y}_{m_0}$ as the following Young diagram with m_0 boxes:

$$\tilde{\boldsymbol{\mu}} := (\tilde{\mu}_1, \ldots, \tilde{\mu}_d) \quad \text{s.t.} \quad \sum_i |\tilde{\mu}_i| = m_0 \quad \text{and}$$

$$\tilde{\mu}_j + 1 \geqslant \mu_i \geqslant \tilde{\mu}_j \quad \forall j > i. \tag{B5}$$

Now we define the following viable subset of Young diagrams with d rows and m boxes, on which our probe state has support:

$$\mathsf{S}_{\text{via}} := \{\boldsymbol{\lambda} \in \mathsf{Y}_m : \exists \tilde{\boldsymbol{\lambda}} \in [M]^{\times(d-1)} \text{ s.t.}$$

$$\lambda_i = \tilde{\mu}_i + (2d - i - 2)M + d - i + \tilde{\lambda}_i,$$

$$i = 1, \ldots, d - 1\}. \tag{B6}$$

The reference frame state we use in the weak error model is defined by

$$q_{\tilde{\boldsymbol{\lambda}}} := \prod_{i=1}^{d-1} g_{\tilde{\lambda}_i}, \tag{B7}$$

where g is the following distribution over $[M]$:

$$g_{\tilde{\lambda}_i} := \frac{2}{M+1}\sin^2\left(\frac{\pi(2\tilde{\lambda}_i + 1)}{2(M+1)}\right). \tag{B8}$$

A similar construction has recently been used to achieve optimal programming of unitary gates [69].

2. Bounding the error

In the current error model, the code $(\mathcal{E}, \mathcal{D})$ can be made exact (i.e. error-free). Therefore, to use Lemma 1, we simply need to bound $\sum_{\lambda \in \mathsf{S}_{\text{int}}} \sqrt{q_\lambda q_{\lambda+\Delta}}$ and appeal to Lemma 2. In the following, we show that our choice of $\{q_\lambda\}$ [see Eq. (B7)] satisfies

$$\min_{\Delta \in \mathsf{S}_{\text{diff}}} \sum_{\lambda \in \mathsf{S}_{\text{int}}} \sqrt{q_\lambda q_{\lambda+\Delta}} \geqslant 1 - \epsilon \tag{B9}$$

for $\epsilon = \frac{d}{2}(\frac{\pi n'}{M+1})^2 + O(M^{-3})$. In this case, we can define S_{int} as given by Eq. (B6) with the additional constraint that $2n' \leqslant \tilde{\lambda}_i \leqslant M - 2n'$ for every i. First, denote by ϵ_g the quantity

$$\epsilon_g(\delta) := 1 - \sum_{k=2n'}^{M-2n'} \sqrt{g_k g_{k+\delta}} \leqslant \frac{1}{2}\left(\frac{\pi\delta}{M+1}\right)^2 + O(M^{-3}). \tag{B10}$$

The inequality can be shown by straightforward calculation (see Subsection B 4). For this distribution, it is straightforward that

$$\sum_{\boldsymbol{\lambda} \in S_{\text{int}}} \sqrt{q_{\boldsymbol{\lambda}} q_{\boldsymbol{\lambda}+\boldsymbol{\Delta}}} \geqslant 1 - \frac{d}{2}\left(\frac{\pi n'}{M+1}\right)^2 - O(M^{-3}). \quad \text{(B11)}$$

for any $\boldsymbol{\Delta} \in S_{\text{diff}}$. Summarizing, we reach the bound

$$F_{\text{wc}}\left(\int dU \, p(U|I)\mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{L}}^*, \mathcal{I}_{\text{C}} \otimes \mathcal{I}_{\text{L}}\right)$$
$$\geqslant 1 - \frac{d}{2}\left(\frac{\pi n'}{M+1}\right)^2 - O(M^{-3}). \quad \text{(B12)}$$

Substituting Eqs. (B3), (B4), and $n' = n_{\text{C}} + d - 1$ into the above bound, we get:

$$F_{\text{wc}} \geqslant 1 - \frac{9\pi^2 d^3 (d-1)^2 (n_{\text{e}}+1)^2 (n_{\text{C}}+d-1)^2}{2n_{\text{R}}^2} - O\left(n_{\text{R}}^{-3}\right). \quad \text{(B13)}$$

Applying Lemma 1 and recalling that the code error is always zero, we obtain the performance of our protocol:

Theorem 1 (Heisenberg-limited covariant error correction). For the weak error model, defined by Eq. (B1), the diamond norm error of Protocol 1 is upper bounded as

$$\epsilon_{\text{cov}} \leqslant \frac{81\pi^2 d^4 (d-1)^2 (n_{\text{e}}+1)^2 (n_{\text{C}}+d-1)^2}{2n_{\text{R}}^2} + O\left(n_{\text{R}}^{-3}\right). \quad \text{(B14)}$$

The reference frame register in Protocol 1 should be initiated in the state $\psi^{\otimes(n_{\text{e}}+1)}$, where the state ψ is prepared in the form Eq. (A9) with coefficients given by Eq. (B7).

Since $n_{\text{R}} = n - n_{\text{C}}$ and n_{C} can be chosen to be $O(1)$, our protocol achieves the Heisenberg limit $1/n^2$ with respect to the total number of qudit systems.

3. Optimality of Protocol 1 under the weak error model

Here we prove the optimality of our protocol under the weak error model Eq. (B1). In particular, we consider any code constructed on n qudit systems, denoted as $(\mathcal{E}_{\text{cov}}, \mathcal{D}_{\text{cov}})$, that is covariant under the $\mathsf{SU}(d)$ action. For the weak error model, we show that the Heisenberg limit $1/n^2$ is the ultimate limit for any covariant code, when each qudit has an equal probability $1/n$ of being erased. We stress the full generality of this result, in the sense that it does not assume any specific structure of the code.

The optimality can be shown by applying the lower bound on ϵ_{cov}, as given in Lemma 4, to the weak error model Eq. (B1). We focus on the case when either exactly n_{e} qudits are erased or no qudit is erased at all, i.e., $p_{\mathbf{s}} = 0$ for $0 < |\mathbf{s}| < n_{\text{e}}$. Notice that Ref. [47], where the original lower bound was derived, considered only independent local errors, which is not the case here. However, as long as we can show $I_{\text{Fisher}}^{\uparrow} < \infty$, the bound will work for our model. In the following we identify a Kraus form as in Eq. (A55). Consider the following Kraus form of $(\mathcal{C}_{\text{e}})_{\mathbf{s}}$, the erasure of qudits with labels in $\mathbf{s} \subset [n]$, that depends on θ:

$$(\mathcal{C}_{\text{e}})_{\mathbf{s}}(\cdot) = \sum_{\boldsymbol{n}} C_{\mathbf{s},\boldsymbol{n};\theta}(\cdot)C_{\mathbf{s},\boldsymbol{n};\theta}^{\dagger}$$

$$C_{\boldsymbol{n},\mathbf{s};\theta} := \prod_{j=1}^{n_{\text{e}}}\left(\exp\left\{\frac{i\theta h_{n_j}}{\binom{n-1}{n_{\text{e}}-1}p_{\mathbf{s}}}\right\}|d+1\rangle\langle j|_{\mathbf{s}_j}\right).$$

Here, for convenience, we add a state $|d+1\rangle$ as the state after erasure, $\boldsymbol{n} = (n_1, \ldots, n_{n_{\text{e}}}) \in (\mathbb{Z}_{d+1})^{\otimes n_{\text{e}}}$, $H = \sum_{j=1}^{d+1} h_j|j\rangle\langle j|$ with $h_{d+1} := 0$, $A_l := I_1 \otimes \cdots \otimes I_{l-1} \otimes A \otimes I_{l+1} \otimes \cdots \otimes I_n$ for any operator A, and $\mathbf{s}_j$ refers to the j-th largest element of $\mathbf{s}$.

The entire channel is $\mathcal{C}_{\theta} := \mathcal{C} \circ \mathcal{U}_{\theta}^{\otimes n}$ with $\mathcal{C}$ defined by Eq. (B1), which has Kraus operators

$$K_{0;\theta} = \sqrt{p_0}\bigotimes_{l=1}^{n} U_{l;\theta}$$

$$K_{\boldsymbol{n},\mathbf{s};\theta} = \sqrt{p_{\mathbf{s}}}\prod_{j=1}^{n_{\text{e}}}\left(\exp\left\{\frac{i\theta h_{n_j}}{\binom{n-1}{n_{\text{e}}-1}p_{\mathbf{s}}}\right\}|d+1\rangle\langle n_j|_{\mathbf{s}_j}\right)\left(\bigotimes_{l=1}^{n} U_{l;\theta}\right)$$

$$\mathbf{s} \subset [n], \ |\mathbf{s}| = n_{\text{e}} \quad \boldsymbol{n} \in (\mathbb{Z}_{d+1})^{\otimes n_{\text{e}}}.$$

Their derivatives are

$$\dot{K}_{0;\theta} = i\sqrt{p_0}\left(\bigotimes_l U_{l;\theta}\right)\left(-\sum_l H_l\right)$$

$$\dot{K}_{\boldsymbol{n},\mathbf{s};\theta} = i\sqrt{p_{\mathbf{s}}}\prod_{j=1}^{n_{\text{e}}}\left(\exp\left\{\frac{i\theta h_{n_j}}{\binom{n-1}{n_{\text{e}}-1}p_{\mathbf{s}}}\right\}|d+1\rangle\langle n_j|_{\mathbf{s}_j}\right)\left(\bigotimes_l U_{l;\theta}\right)\left(\sum_{j=1}^{n_{\text{e}}} \frac{h_{n_j}}{\binom{n-1}{n_{\text{e}}-1}p_{\mathbf{s}}}|n_j\rangle\langle n_j|_{\mathbf{s}_j} - \sum_l H_l\right).$$

where $l = 1, \ldots, n$. One can verify that

$$\dot{K}_{0;\theta}^{\dagger}K_{0;\theta} + \sum_{\mathbf{s}}\sum_{\boldsymbol{n}} \dot{K}_{\boldsymbol{n},\mathbf{s};\theta}^{\dagger}K_{\boldsymbol{n},\mathbf{s};\theta}$$

$$= ip_0\left(\sum_l H_l\right) - i\sum_{\mathbf{s}} p_{\mathbf{s}}\left(\sum_{\boldsymbol{n}}\left(\sum_j \frac{h_{n_j}}{\binom{n-1}{n_{\text{e}}-1}p_{\mathbf{s}}}|n_j\rangle\langle n_j|_{\mathbf{s}_j} - \sum_l H_l\right)\left(\prod_j |n_j\rangle\langle n_j|_{\mathbf{s}_j}\right)\right)$$

$$= i p_0 \left(\sum_l H_l \right) - i \left(\sum_{\mathsf{s}} p_{\mathsf{s}} \sum_j \frac{H_{\mathsf{s}_j}}{\binom{n-1}{n_{\mathrm{e}}-1} p_{\mathsf{s}}} - (1 - p_0) \sum_l H_l \right)$$

$$= 0. \tag{B15}$$

In the meantime

$$\dot{K}_{0;\theta}^{\dagger} \dot{K}_{0;\theta} + \sum_{\mathsf{s}} \sum_{\boldsymbol{n}} \dot{K}_{\boldsymbol{n},\mathsf{s};\theta}^{\dagger} \dot{K}_{\boldsymbol{n},\mathsf{s};\theta}$$

$$= \sum_{\mathsf{s}} \sum_{j=1}^{n_{\mathrm{e}}} \frac{(H^2)_{\mathsf{s}_j}}{\binom{n-1}{n_{\mathrm{e}}-1}^2 p_{\mathsf{s}}} - \left(\sum_l H_l \right)^2. \tag{B16}$$

Therefore the Fisher information bound satisfies

$$I_{\mathrm{Fisher}}^{\uparrow} \leqslant 4 \left\| \dot{K}_{0;\theta}^{\dagger} \dot{K}_{0;\theta} + \sum_{\mathsf{s}} \sum_{\boldsymbol{n}} \dot{K}_{\boldsymbol{n},\mathsf{s};\theta}^{\dagger} \dot{K}_{\boldsymbol{n},\mathsf{s};\theta} \right\|_{\infty}$$

$$\leqslant 4 \sum_{\mathsf{s},j} \frac{1}{\binom{n-1}{n_{\mathrm{e}}-1}^2 p_{\mathsf{s}}} \|H^2\|_{\infty} + 4n^2 \|H\|_{\infty}^2$$

$$= 4 \sum_{\mathsf{s}} \frac{n_{\mathrm{e}}}{\binom{n-1}{n_{\mathrm{e}}-1}^2 p_{\mathsf{s}}} \|H^2\|_{\infty} + 4n^2 \|H\|_{\infty}^2. \tag{B17}$$

Substituting into Eq. (A56), we get

$$\epsilon_{\mathrm{wc}} \geqslant \frac{(\Delta H)^2}{64 \left(\sum_{\mathsf{s}} n_{\mathrm{e}} \|H^2\|_{\infty} / (p_{\mathsf{s}} \binom{n-1}{n_{\mathrm{e}}-1}^2) + n^2 \|H\|_{\infty}^2 \right)}. \tag{B18}$$

Now, we assume that each qudit has an equal probability $p_{\mathsf{s}} = 1/\binom{n}{n_{\mathrm{e}}}$ of being erased and $p_0 = 0$. Since $\|H^2\|_{\infty} =$

$\|H\|_{\infty}^2$, the above bound implies

$$\epsilon_{\mathrm{cov}} \geqslant \frac{(\Delta H)^2}{64 n^2 (1 + 1/n_{\mathrm{e}}) \|H\|_{\infty}^2}. \tag{B19}$$

We can choose the minimum eigenvalue and the maximum eigenvalue of H to sum up to zero. Then, $\Delta H = 2\|H\|_{\infty}$, and we reach the following proposition:

Proposition 1. For the weak erasure error model Eq. (B1) with $p_{\mathsf{s}} = 1/\binom{n}{n_{\mathrm{e}}}$ for any $\mathsf{s} : |\mathsf{s}| = n_{\mathrm{e}}$, the error of any covariant code is lower bounded as

$$\epsilon_{\mathrm{cov}} \geqslant \frac{1}{16 n^2 (1 + 1/n_{\mathrm{e}})}. \tag{B20}$$

Since the above bound matches the performance of our protocol in scaling, we conclude that the optimal error scaling of covariant codes is identified as $1/n^2$.

4. Proof of Eq. (B10)

Invoking the definition of $\{g_k\}$ from Eq. (B7), we have the following chain of (in)equalities:

$$\sum_{k=n}^{M-n} \sqrt{g_k g_{k+\delta}} = \frac{2}{M+1} \sum_{k=n}^{M-n} \sin \left(\frac{\pi(2k+1)}{2(M+1)} \right) \sin \left(\frac{\pi(2k+2\delta+1)}{2(M+1)} \right)$$

$$= \frac{1}{M+1} \sum_{k=n}^{M-n} \left(\cos \left(\frac{\pi\delta}{M+1} \right) - \cos \left(\frac{\pi(2k+\delta+1)}{M+1} \right) \right)$$

$$= \frac{1}{M+1} \left((M-2n+1) \cos \left(\frac{\pi\delta}{M+1} \right) - \sum_{k=n}^{M-n} \cos \left(\frac{\pi(2k+\delta+1)}{M+1} \right) \right)$$

$$= \frac{1}{M+1} \left(\left(M-2n+1 + \frac{\sin \left(\frac{2\pi n}{M+1} \right)}{\sin \left(\frac{\pi}{M+1} \right)} \right) \cos \left(\frac{\pi\delta}{M+1} \right) \right)$$

$$= \frac{\cos \left(\frac{\pi\delta}{M+1} \right)}{M+1} \left(M-2n+1 + \cos \left(\frac{\pi}{M+1} \right) + \sum_{k=1}^{2n-1} \cos \left(\frac{\pi k}{M+1} \right) \right)$$

$$\geqslant \frac{1}{M+1} \left(M+1 - \frac{1}{2} \left(\frac{\pi}{M+1} \right)^2 - \sum_{k=1}^{2n-1} \frac{1}{2} \left(\frac{\pi k}{M+1} \right)^2 \right) \left(1 - \frac{1}{2} \left(\frac{\pi\delta}{M+1} \right)^2 \right)$$

$$= 1 - \frac{1}{2} \left(\frac{\pi\delta}{M+1} \right)^2 - O(M^{-3}).$$

APPENDIX C: THE i.i.d. ERROR MODEL

1. Setting

In this section, we deal with another type of erasure errors which are stronger than the one considered in the previous section. The error affects each qudit independently, erasing it with a probability:

$$\mathcal{C} = \bigotimes_{j=1}^{n} ((1 - p_{\text{e}})\mathcal{I} + p_{\text{e}}\mathcal{C}_{\text{e}}) \quad p_{\text{e}} \in \left(0, \frac{1}{2}\right), \qquad \text{(C1)}$$

where $\mathcal{C}_{\text{e}}$ denotes the single-qudit erasure channel. Since the erasure channel is degradable, if $p_{\text{e}} \geqslant 1/2$ the information leaked to the environment would not be retrievable. Note that, in general, the error model on each qudit does not have to be identical, and each qudit j can have distinct probability $p_{\text{e},j}$ of being erased. In that case, however, we can simply set p_{e} to be the worst case over $\{p_{\text{e},j}\}$ and consider this more stringent model instead. We can cast Eq. (C1) in the form of Eq. (A17) as

$$\mathcal{C} = \sum_{\mathbf{s} \subset [n]} p_{\text{e},\mathbf{s}} \left(\bigotimes_{k \in \mathbf{s}} \mathcal{C}_{\text{e}} \right) \otimes \left(\bigotimes_{k' \in \mathbf{s}^c} \mathcal{I} \right) \qquad \text{(C2)}$$

where $p_{\text{e},\mathbf{s}} := p_{\text{e}}^{|\mathbf{s}|} (1 - p_{\text{e}})^{n - |\mathbf{s}|}$.

In contrast to the previous model, there does not exist any code that perfectly corrects independent erasure errors. Instead, there exist pretty good codes that correct the error unless too many qudits are erased. The quantum capacity of the erasure channel, with erasure probability p_{e}, has been determined to be $1 - 2p_{\text{e}}$ [49]. We can choose n_{C}, the number of computational qudits, to grow with n. When n is large, since the number of qudits we want to encode is only one qudit and is much smaller than that allowed by the capacity (which is $(1 - 2p_{\text{e}})n_{\text{C}}$), the error probability would vanish exponentially in n_{C}.

We choose $n_{\text{C}} = n^{\gamma}$, where $\gamma \in (0, 1)$ does not depend on n and can be chosen to be very small. We use any code $(\mathcal{E}, \mathcal{D})$ in Protocol 1 that encodes one qudit into n^{γ} computational qudits, with the property that it has a decoding error $O(e^{-x_d \cdot n^{\gamma}})$ for some $x_d > 0$ that may depend on d. By a random coding argument, one can show that there exists a stabilizer code satisfying our requirement (see, e.g., [50]), although its explicit form is not given. Recently progress in error correcting codes also showed that quantum polar codes [51,52] and Reed-Muller codes [53] have the desired property. Notice that the requirement of the $O(e^{-x_d \cdot n^{\gamma}})$ scaling is chosen for convenience of analyzing the error, and it can be further relaxed in practice.

Meanwhile, the model is now too noisy for the highly coupled reference frame state used in the weak error model to be effective. Instead, we prepare on the reference frame register the following s_{R}-copy state:

$$\Psi = (\Phi^+)^{\otimes s_{\text{R}}} \qquad \text{(C3)}$$

with $s_{\text{R}} = n_{\text{R}}/2 = (1 - n^{-1+\gamma})n/2$ and $|\Phi^+\rangle$ being the maximally entangled state on $\mathcal{H} \otimes \mathcal{H}$. Intuitively, this choice is to distribute the eggs in different baskets. One erasure error destroys at most one of the s_{R} reference frame states. As long as there are still $\Theta(n)$ copies left we can achieve high

performance, which happens with very high probability since $p_{\text{e},\mathbf{s}}$ is a binomial distribution.

2. Bounding the error

When s_{e} reference frames are erased, the number of remaining reference frames is

$$s := s_{\text{R}} - s_{\text{e}}. \qquad \text{(C4)}$$

The remaining reference frames can be decomposed in the form Eq. (A9) as

$$(\Phi^+)^{\otimes s} = \bigoplus_{\lambda \in Y_s} \sqrt{p_{\lambda,s}} |\Phi_\lambda^+\rangle \otimes |\Phi_{m_\lambda}^+\rangle \qquad \text{(C5)}$$

where $p_{\lambda,s}$ is known as the Schur-Weyl distribution. It has the following precise form ([89], Eq. (3.28)):

$$p_{\lambda,s} = \frac{s! \prod_{i<j} (\tilde{\lambda}_i - \tilde{\lambda}_j)^2}{d^s (\tilde{\lambda}_d)! \prod_{k=1}^{d-1} k! (\tilde{\lambda}_k)!} \quad \tilde{\lambda}_j := \lambda_j + d - j. \qquad \text{(C6)}$$

We now combine it with Eq. (A12) and Lemma 3 to express the fidelity Eq. (A36) as

$$F_s = \sum_{\lambda, \lambda' \in Y_s} \sqrt{p_{\lambda,s} p_{\lambda',s}} S_{\lambda,\lambda'}, \qquad \text{(C7)}$$

where $S_{\lambda,\lambda'}$ is the correlation function defined by Eq. (A43). For any $\alpha > 0$, if $s = \beta \cdot n$ for some fixed $\beta > 0$, then there exists $n_{0,\alpha}$ so that

$$1 - F_s \leqslant \left(\frac{(d^2 - d + 32)\Gamma}{\beta^2} \right) \cdot \left(\frac{1}{2s} \right)^{1 - \alpha - 2\gamma} \qquad \text{(C8)}$$

holds for arbitrary $n \geqslant n_{0,\alpha}$. Here $\Gamma := d^{\frac{d(d-1)}{2}} / (\prod_{k=1}^{d-1} k!)$.

Eq. (C8) implies that $1 - F_s$ scales almost as $1/s$ for large s. Its proof is delayed to Subsection C 4. Roughly speaking, in the large s limit, $p_{\lambda,s}$ converges to a tilted multi-variate Gaussian. We can focus on a region near the peak of $p_{\lambda,s}$ where the distribution is concentrated and where the correlation function $S_{\lambda,\lambda'}$ is maximized as well. This would introduce at most an error vanishing as $1/s$. In addition, since a Gaussian is relatively flat around its peak, we can show that $\sqrt{p_{\lambda+\Delta,s} p_{\lambda,s}} \approx p_{\lambda+\Delta/2,s}$ up to an error that scales as $1/s$. This would be the main reason of getting this error scaling.

With Eq. (C8) it is rather straightforward to bound the error of our protocol. First we define a threshold value

$$s^* := (1 - 2p_{\text{e}}) \cdot s_{\text{R}} = (1/2 - p_{\text{e}})(n - n^{\gamma}). \qquad \text{(C9)}$$

This definition ensures that the probability that $s < s^*$ goes to zero exponentially fast. Next, notice that the error probabilities on the computational register and the reference frame register are independent:

$$p_{\text{e}}(\mathbf{s}) = p_{\text{e,P}}(\mathbf{s}_C) p_{\text{e,R}}(\mathbf{s}_R) \quad \mathbf{s} = \mathbf{s}_C \cup \mathbf{s}_R, \qquad \text{(C10)}$$

where $p_{\text{e,P}}$ and $p_{\text{e,R}}$ are the probability distributions of erasure errors in the computational register and in the reference frame register, respectively.

Applying Lemma 1 and Eq. (C10), we can split the error of our protocol as

$$
\epsilon_{\text{cov}} \leqslant 9d \left(\sum_{\mathsf{s} \subset [n], F_s \geqslant \frac{3}{4}} p_{\text{e}}(\mathsf{s}) \max\{\epsilon_{\mathsf{s}_{\text{C}}, \text{code}}, 1 - F_s\} + \sum_{\mathsf{s} \subset [n], F_s < \frac{3}{4}} p_{\text{e}}(\mathsf{s}) \right). \tag{C11}
$$

From Eqs. (C8) and (C9), we know that, for large enough n, $F_s > \frac{3}{4}$ for $s \geqslant s^*$. That is, when no more than $2p_{\text{e}}s_{\text{R}}$ erasure errors occur on the reference frame register. We can then express the bound as

$$
\epsilon_{\text{cov}} \leqslant 9d \left(\sum_{\mathsf{s}_{\text{C}}} p_{\text{e},\text{P}}(\mathsf{s}_{\text{C}}) \epsilon_{\mathsf{s}_{\text{C}}, \text{code}} + \sum_{\mathsf{s}_{\text{R}}} p_{\text{e},\text{R}}(\mathsf{s}_{\text{R}})(1 - F_s) + \sum_{\mathsf{s}_{\text{R}}: |\mathsf{s}_{\text{R}}| > 2p_{\text{e}}s_{\text{R}}} p_{\text{e},\text{R}}(\mathsf{s}_{\text{R}}) \right). \tag{C12}
$$

Notice that, by our assumption on $(\mathcal{E}, \mathcal{D})$, we should have

$$
\sum_{\mathsf{s}_{\text{C}}} p_{\text{e},\text{P}}(\mathsf{s}_{\text{C}}) \epsilon_{\mathsf{s}_{\text{C}}, \text{code}} = O(e^{-x_d \cdot n^{\gamma}}). \tag{C13}
$$

Therefore the reference frame error constitutes the main contribution to the overall error:

$$
\begin{aligned}
\epsilon_{\text{cov}} &\leqslant 9d \left(\sum_{\mathsf{s}_{\text{R}}} p_{\text{e},\text{R}}(\mathsf{s}_{\text{R}})(1 - F_s) + O(e^{-x_d \cdot n^{\gamma}}) + \sum_{\mathsf{s}_{\text{R}}: |\mathsf{s}_{\text{R}}| > 2p_{\text{e}}s_{\text{R}}} p_{\text{e},\text{R}}(\mathsf{s}_{\text{R}}) \right) \quad s = \frac{n_{\text{R}}}{2} - |\mathsf{s}_{\text{R}}| \\
&\leqslant 9d \left(1 - F_{s^*} + 2 \sum_{\mathsf{s}_{\text{R}}: |\mathsf{s}_{\text{R}}| > 2p_{\text{e}}s_{\text{R}}} p_{\text{e},\text{R}}(\mathsf{s}_{\text{R}}) \right) + O(e^{-x_d \cdot n^{\gamma}}) \\
&\leqslant \left(\frac{36d(d^2 - d + 32)\Gamma}{(1 - 2p_{\text{e}})^2} \right) \cdot \left(\frac{1}{(1 - 2p_{\text{e}})n} \right)^{1 - \alpha - 2\gamma} + O(e^{-p_{\text{e}}^2(1 - \gamma)n}) + O(e^{-x_d \cdot n^{\gamma}}). \tag{C14}
\end{aligned}
$$

The second inequality comes from dividing the summation into the term for $s \geqslant s^*$ and the term for $s < s^*$. The third inequality comes from Eq. (C8) (with $\beta = 1/2 - p_{\text{e}}$) and the Hoeffding bound. From the above bound, we can see that γ can be chosen to be arbitrarily close to zero. Absorbing the other error terms into the major term, we have the following bound on the error:

Theorem 2. Consider the i.i.d. error model defined by Eq. (C1). For any $\alpha > 0$, there exists $n_{\alpha} > 0$ so that the diamond norm error of Protocol 1 is upper bounded by

$$
\epsilon_{\text{cov}} \leqslant \left(\frac{36(d^2 - d + 32)d^{\frac{d^2 - d + 2}{2}}}{(1 - 2p_{\text{e}})^2 \prod_{j=1}^{d}(j-1)!} \right) \left(\frac{1}{(1 - 2p_{\text{e}})n} \right)^{1 - \alpha} \tag{C15}
$$

for any $n \geqslant n_{\alpha}$. The reference frame register in Protocol 1 should be initiated in multiple copies of the maximally entangled qudit state.

The error of our protocol scales almost as $1/n$, instead of $1/n^2$ in the previous case. This is a result of the (stronger) i.i.d. noise. In fact, as we show in the next section, the error scaling of our protocol is still optimal for this error model.

3. Optimality of Protocol 1 under the i.i.d. error model

Here we prove the optimality of our protocol under the i.i.d. error model Eq. (C1). In particular, we consider any code constructed on n qudit systems, denoted as $(\mathcal{E}_{\text{cov}}, \mathcal{D}_{\text{cov}})$, that is covariant under the $\mathsf{SU}(d)$ action. For the i.i.d. error model, we show that the $1/n$ scaling is optimal.

Under the i.i.d. error model Eq. (C1), the optimality of our protocol can, again, be shown by applying Lemma 4. Here the

Fisher information upper bound $I_{\text{Fisher}}^{\uparrow}$ should be replaced by the one for the local, independent erasure model. For the i.i.d. (erasure) error model, the Fisher information upper bound has already been given by ([47], Eq. (B19)) as

$$
I_{\text{Fisher}}^{\uparrow} = 4n(\Delta H)^2 \left(\frac{1 - p_{\text{e}}}{p_{\text{e}}} \right). \tag{C16}
$$

Substituting into Lemma 4, we obtain:

Proposition 2. For the i.i.d. (erasure) error model, defined by Eq. (C1), the error of any covariant code is lower bounded by

$$
\epsilon_{\text{cov}} \geqslant \frac{p_{\text{e}}}{64n(1 - p_{\text{e}})}. \tag{C17}
$$

Therefore, since Protocol 1 achieves the $(1/n)$ scaling, it is optimal in the asymptotic limit of large n.

4. Proof of Eq. (C8)

Here we prove Eq. (C8) under the assumption that $s = \beta \cdot n$ for some fixed $\beta > 0$. Since $S_{\lambda, \lambda'} \geqslant 0$ we can lower bound the fidelity Eq. (C7) as

$$
F_s \geqslant \sum_{\lambda, \lambda' \in \mathsf{S}_{\text{cent}}} \sqrt{p_{\lambda, s} p_{\lambda', s}} S_{\lambda, \lambda'}, \tag{C18}
$$

where $\mathsf{S}_{\text{cent}} \subset \mathsf{Y}_s$ is defined as

$$
\mathsf{S}_{\text{cent}} := \left\{ \lambda \in \mathsf{Y}_s : \left| \lambda_i - \frac{s}{d} \right| \leqslant \frac{s^{\frac{1+x}{2}}}{2}, |\lambda_i - \lambda_j| > 3n', \forall i, j \right\} \tag{C19}
$$

$$
n' = n^{\gamma} + d - 1. \tag{C20}
$$

Here $x > 0$ is a parameter to be specified at the end of the proof. We can express the fidelity as

$$F_s \geqslant \sum_{\lambda \in S_{\text{cent}}} \sum_{\Delta \in S_{\text{diff}}} \sqrt{p_{\lambda,s} p_{\lambda+\Delta,s}} S_{\lambda,\lambda+\Delta}. \quad (C21)$$

where

$$S_{\text{diff}} = \{\Delta \in \mathbb{Z}^{\times d} : |\Delta_i| \leqslant n', i = 1, \ldots, d; \sum_{j=1}^{d} \Delta_j = 0\}. \quad (C22)$$

Similar as in the proof of Lemma 2, we first show that: For $\lambda \in S_{\text{cent}}$ and $\Delta \in S_{\text{diff}}$, the correlation function depends only on their relative distance Δ, but not explicitly on λ, i.e., $S_{\lambda,\lambda+\Delta} = \tilde{S}_\Delta$. Indeed, when λ is in the viable set S_{cent}, the lengths of different rows of both λ and $\lambda + \Delta$ have big enough gaps so that adding n' boxes to one row would not make its box number greater than its preceding rows. Therefore, adding μ' and μ in S_{cost} according to the Littlewood-Richardson rule is not constraint by the shape of λ. Now, the fidelity bound becomes

$$F_s \geqslant \sum_{\lambda \in S_{\text{cent}}} \sum_{\Delta \in S_{\text{diff}}} \sqrt{p_{\lambda,s} p_{\lambda+\Delta,s}} \tilde{S}_\Delta \quad (C23)$$

$$\geqslant \left(\min_{\Delta \in S_{\text{diff}}} \sum_{\lambda \in S_{\text{cent}}} \sqrt{p_{\lambda,s} p_{\lambda+\Delta,s}} \right) \left(\sum_{\Delta' \in S_{\text{diff}}} \tilde{S}_{\Delta'} \right) \quad (C24)$$

$$\geqslant \min_{\Delta \in S_{\text{diff}}} \sum_{\lambda \in S_{\text{cent}}} \sqrt{p_{\lambda,s} p_{\lambda+\Delta,s}}, \quad (C25)$$

having used Eq. (A51) in the last step.

What remains is to bound $\sum_{\lambda \in S_{\text{cent}}} \sqrt{p_{\lambda,s} p_{\lambda+\Delta,s}}$ for every $\Delta \in S_{\text{diff}}$, where $p_{\lambda,s}$ is the Schur-Weyl distribution:

$$p_{\lambda,s} = \frac{s! \prod_{i<j}(\tilde{\lambda}_i - \tilde{\lambda}_j)^2}{d^s (\tilde{\lambda}_d)! \prod_{k=1}^{d-1} k! (\tilde{\lambda}_k)!} \quad \tilde{\lambda}_j := \lambda_j + d - j. \quad (C26)$$

First, we show that for the Schur-Weyl distribution $p_{\lambda,s}$, $\sqrt{p_{\lambda+\Delta,s} p_{\lambda,s}} \approx p_{\lambda+\Delta/2,s}$ for $\lambda \in S_{\text{cent}}$. Define the following multinomial distribution of $\tilde{\lambda}$:

$$b_{\tilde{\lambda}} := \frac{(s + d(d-1)/2)!}{(\tilde{\lambda}_1)! \cdots (\tilde{\lambda}_d)!} d^{-(s+d(d-1)/2)}, \quad (C27)$$

and we can bound the Schur-Weyl distribution as

$$p_{\lambda,s} = b_{\tilde{\lambda}} \frac{\Gamma}{[s + d(d-1)/2]_{d(d-1)/2}} \prod_{i<j}(\tilde{\lambda}_i - \tilde{\lambda}_j)^2 \quad (C28)$$

$$\leqslant b_{\tilde{\lambda}} \Gamma \prod_{i<j} \left(\frac{\tilde{\lambda}_i - \tilde{\lambda}_j}{\sqrt{s}} \right)^2, \quad (C29)$$

where $[l]_k := l!/(l-k)!$ and $\Gamma := d^{\frac{d(d-1)}{2}}/(\prod_{k=1}^{d-1} k!)$. We have

$$\frac{\sqrt{p_{\lambda+\Delta,s} p_{\lambda,s}}}{p_{\lambda+\Delta/2,s}}$$

$$= \frac{\sqrt{b_{\tilde{\lambda}+\Delta} b_{\tilde{\lambda}}}}{b_{\tilde{\lambda}+\Delta/2}} \prod_{i<j} \left(1 - \left(\frac{\Delta_i - \Delta_j}{2\tilde{\lambda}_i - 2\tilde{\lambda}_j + \Delta_i - \Delta_j} \right)^2 \right). \quad (C30)$$

Using Stirling's approximation $n! = \sqrt{2\pi n}(n/e)^n (1 + 1/(12n) + O(n^2))$ and $\sum_{i=1}^{d} \Delta_i = 0$, we have

$$\frac{\sqrt{b_{\tilde{\lambda}+\Delta} b_{\tilde{\lambda}}}}{b_{\tilde{\lambda}+\Delta/2}}$$

$$\leqslant \exp\left\{ -\frac{f_{\tilde{\lambda}+\Delta} + f_{\tilde{\lambda}}}{2} + f_{\tilde{\lambda}+\Delta/2} \right\} \left(1 + \frac{d^2}{6s} + O(s^{-2}) \right) \quad (C31)$$

for $\lambda \in S_{\text{cent}}$, where $f_{\tilde{\lambda}} := \sum_i (\tilde{\lambda}_i + 1/2) \ln \tilde{\lambda}_i$. Straightforward calculation shows that

$$\frac{f_{\tilde{\lambda}} + f_{\tilde{\lambda}+\Delta}}{2} - f_{\tilde{\lambda}+\Delta/2} = \sum_i \frac{(\Delta_i)^2}{8\tilde{\lambda}_i} + O\left(\frac{(n')^3}{s^2} \right). \quad (C32)$$

Since $n' = O(n^\gamma)$ and $s = \Theta(n)$, we have

$$\frac{f_{\tilde{\lambda}} + f_{\tilde{\lambda}+\Delta}}{2} - f_{\tilde{\lambda}+\Delta/2} = \sum_i \frac{(\Delta_i)^2}{8\tilde{\lambda}_i} + O(n^{-2+3\gamma}), \quad (C33)$$

which, plus $\sum_i (\Delta_i)^2 \leqslant (n')^2$ and $|\lambda_i - s/d| = O(s^{\frac{1+x}{2}})$, implies that

$$\sqrt{b_{\tilde{\lambda}+\Delta} b_{\tilde{\lambda}}} \geqslant b_{\tilde{\lambda}+\Delta/2} \left(1 - \frac{d(n')^2}{8s} - \frac{d^2}{6s} - O(n^{-\frac{3-x-4\gamma}{2}}) \right) \quad (C34)$$

$$= b_{\tilde{\lambda}+\Delta/2}(1 - O(n^{-1+2\gamma}) - O(n^{-\frac{3-x-4\gamma}{2}})). \quad (C35)$$

Combining the above bound with Eqs. (C22) and (C30), we get

$$\frac{\sqrt{p_{\lambda+\Delta,s} p_{\lambda,s}}}{p_{\lambda+\Delta/2,s}} \geqslant (1 - O(n^{-1+2\gamma}) - O(n^{-\frac{3-x-4\gamma}{2}}))$$

$$\cdot \prod_{i<j} \left(1 - \frac{(n')^2}{(\tilde{\lambda}_i - \tilde{\lambda}_j + \Delta_i/2 - \Delta_j/2)^2} \right)$$

$$\geqslant 1 - \sum_{i<j} \frac{(n')^2}{(\tilde{\lambda}_i - \tilde{\lambda}_j + \Delta_i/2 - \Delta_j/2)^2}$$

$$- O(n^{-1+2\gamma}) - O(n^{-\frac{3-x-4\gamma}{2}}) \quad (C36)$$

and we have $\sqrt{p_{\lambda+\Delta,s} p_{\lambda,s}} \approx p_{\lambda+\Delta/2,s}$ with the approximation error given by Eq. (C36).

What remains is to evaluate the summation $\sum_{\lambda \in S_{\text{cent}}} p_{\lambda+\Delta/2,s}$. Using the bounds Eq. (C19) and (C29), the error terms can be bounded as follows:

$$\frac{\sum_{\lambda \in S_{\text{cent}}} p_{\lambda+\Delta/2,s}}{(\tilde{\lambda}_i - \tilde{\lambda}_j + \Delta_i/2 - \Delta_j/2)^2}$$

$$\leqslant \sum_{\lambda \in S_{\text{cent}}} \frac{\Gamma b_{\tilde{\lambda}+\Delta/2}}{s} \cdot \prod_{k<l, k\neq i, l\neq j} \left(\frac{\tilde{\lambda}_k - \tilde{\lambda}_l + \Delta_k/2 - \Delta_l/2}{\sqrt{s}} \right)^2$$

$$\leqslant \frac{\Gamma}{s} \cdot \prod_{k<l, k\neq i, l\neq j} \left(\frac{s^{\frac{1+x}{2}} + l - k + n'}{\sqrt{s}} \right)^2 \sum_{\lambda \in S_{\text{cent}}} b_{\tilde{\lambda}+\Delta/2}$$

$$\leqslant \frac{\Gamma}{s} \cdot \prod_{k<l, k\neq i, l\neq j} \left(\frac{s^{\frac{1+x}{2}} + l - k + n'}{\sqrt{s}} \right)^2 \cdot 1 \quad (C37)$$

Assuming $(\sqrt{2} - 1)s^{\frac{1+x}{2}} \geqslant d - 1 + n' \geqslant l - k + n'$ for any l, k (which always holds for large enough n), we can simplify

the above bound to

$$\frac{\sum_{\lambda \in \mathsf{S}_{\text{cent}}} p_{\lambda + \Delta/2, s}}{(\tilde{\lambda}_i - \tilde{\lambda}_j + \Delta_i/2 - \Delta_j/2)^2} \leqslant 2\Gamma \left(\frac{1}{2s}\right)^{1 - x \cdot \frac{d^2 - d - 2}{2}}. \quad \text{(C38)}$$

In the same manner, we can show that

$$\sum_{\lambda \in \mathsf{S}'_{\text{cent}} \setminus \mathsf{S}_{\text{cent}}} p_{\lambda + \Delta/2, s} \leqslant 2\Gamma(4n' + d - 1)^2 \left(\frac{1}{2s}\right)^{1 - x \cdot \frac{d^2 - d - 2}{2}} \tag{C39}$$

where

$$\mathsf{S}'_{\text{cent}} := \left\{ \lambda \in \mathsf{S}_s : \left| \lambda_i - \frac{s}{d} \right| \leqslant \frac{s^{\frac{1+x}{2}}}{2}, \; \forall i \right\} \tag{C40}$$

is just S_{cent} with the restriction $|\lambda_i - \lambda_j| > n'$ lifted. Indeed, since for $\lambda \in \mathsf{S}'_{\text{cent}} \setminus \mathsf{S}_{\text{cent}}$ there exists a pair (i, j) such that $|\lambda_i - \lambda_j| \leqslant 3n'$ and $|\lambda_k - \lambda_l| \leqslant s^{\frac{1+x}{2}}$ for the rest, we have

$$\sum_{\lambda \in \mathsf{S}'_{\text{cent}} \setminus \mathsf{S}_{\text{cent}}} p_{\lambda + \Delta/2, s}$$

$$\leqslant \sum_{\lambda \in \mathsf{S}'_{\text{cent}} \setminus \mathsf{S}_{\text{cent}}} \Gamma \, b_{\tilde{\lambda} + \Delta/2} \cdot \prod_{k < l} \left(\frac{\tilde{\lambda}_k - \tilde{\lambda}_l + \Delta_k/2 - \Delta_l/2}{\sqrt{s}} \right)^2$$

$$\leqslant \sum_{\lambda \in \mathsf{S}'_{\text{cent}} \setminus \mathsf{S}_{\text{cent}}} \Gamma \, b_{\tilde{\lambda} + \Delta/2} \cdot \left(\frac{3n' + d - 1 + n'}{\sqrt{s}} \right)^2 \cdot (2s)^{x \cdot \frac{d^2 - d - 2}{2}}$$

$$\leqslant 2\Gamma(4n' + d - 1)^2 \cdot \left(\frac{1}{2s} \right)^{1 - x \cdot \frac{d^2 - d - 2}{2}}.$$

The probability that $\lambda \notin \mathsf{S}'_{\text{cent}}$ vanishes exponentially due to large deviation theory. Precisely, the following bound holds [90,91]:

$$\sum_{\lambda \in \mathsf{S}'_{\text{cent}}} p_{\lambda + \Delta/2, s}$$

$$\geqslant 1 - \max_{\lambda' - \Delta/2 \notin \mathsf{S}_{\text{cent}}} (s + 1)^{\frac{d(d-1)}{2}} \exp \left\{ -\frac{2d_{\text{Young}}(\lambda', s/d)^2}{s} \right\} \tag{C41}$$

where s/d here refers to the Young diagram $(s/d, \dots, s/d)$. Substituting in Eq. (C40) and using $s = \Theta(n)$ we get

$$\sum_{\lambda \in \mathsf{S}'_{\text{cent}}} p_{\lambda + \Delta/2, s} \geqslant 1 - O(e^{-n^x}). \tag{C42}$$

Finally, summarizing Eqs. (C36), (C38), (C39), and (C42), we get

$$1 - F_s \leqslant 1 - \min_{\Delta \in \mathsf{S}_{\text{diff}}} \sum_{\lambda \in \mathsf{S}_{\text{cent}}} \sqrt{p_{\lambda, s} p_{\lambda + \Delta, s}} \tag{C43}$$

$$\leqslant \left(\frac{d(d-1)(n')^2}{2} + (4n' + d - 1)^2 \right) 2\Gamma \cdot \left(\frac{1}{2s} \right)^{1 - x \cdot \frac{d^2 - d - 2}{2}}$$

$$+ O(n^{-1+2\gamma}) + O\left(n^{-\frac{3-x-4\gamma}{2}} \right) + O(e^{-n^x}). \tag{C44}$$

Picking any $x \in (0, \min\{2\alpha/(d^2 - d - 2), 1\})$ we get

$$1 - F_s \leqslant \Gamma(d^2 - d + 32) \cdot (n')^2 \left(\frac{1}{2s} \right)^{1-\alpha} + O(n^{-1+2\gamma}). \tag{C45}$$

Recalling that $n' = n^\gamma + d - 1$ and $s = \beta \cdot n$, we get Eq. (C8). Since the above bound holds for any $x > 0$, the scaling of $1 - F_s$ approaches $s^{-1+2\gamma}$ in the asymptotic limit of large s.

APPENDIX D: RESOURCE REQUIREMENTS AND IMPLEMENTATION

1. Compression of quantum reference frames

In both models, the reference frame states are constructed on n_{R} qudits, which is a $(d^{n_{\text{R}}})$-dimensional system. However, here we show that the reference frame states can be compressed into a much smaller system. Indeed, the dimension of the state is given by the effective dimension

$$d_{\text{R}} := \dim \left(\mathsf{Span}\{\mathcal{U}_{\text{R}}(\Psi)\}_{U \in \mathsf{SU}(d)} \right), \tag{D1}$$

which, as we show in the following, grows only polynomially in n_{R}.

Since we always use entangled reference frames, $U_{\text{R}} = (U \otimes I)^{\otimes \frac{n_{\text{R}}}{2}}$. By Schur-Weyl duality, we have the decomposition

$$U_{\text{R}} \simeq \bigoplus_{\lambda \in \mathsf{Y}_{n_{\text{R}}/2}} (U_\lambda \otimes I_{m_\lambda}) \otimes I^{\otimes \frac{n_{\text{R}}}{2}}. \tag{D2}$$

Therefore d_{R}, which is equal to the rank of the twirled state $\int dU \, \mathcal{U}_{\text{R}}(\Psi)$, satisfies the upper bound

$$d_{\text{R}} \leqslant \sum_{\lambda \in \mathsf{Y}_{n_{\text{R}}/2}} d_\lambda^2$$

$$\leqslant |\mathsf{Y}_{n_{\text{R}}/2}| \max_{\lambda \in \mathsf{Y}_{n_{\text{R}}/2}} d_\lambda^2$$

$$\leqslant \left(\frac{n_{\text{R}}}{2} + 1 \right)^{(d^2 - 1)}. \tag{D3}$$

The last bound is quite straightforward to obtain; see, for instance, Ref. ([92], Eqs. (6.16) and (6.18)). Therefore we can compress the reference frame state to a system of much smaller dimension, which reduces exponentially the cost of quantum memory during idle time. The cost can be further reduced at the price of a small recovery error using the compression protocols in Refs. [56–58].

2. Computational efficiency of implementation

At first sight, our encoding and decoding may appear to necessitate sampling from the Haar measure on $\mathsf{SU}(d)$ in order to be implemented [see Eqs. (3) and (4)]. From a computational complexity standpoint, this would be a problem, since to implement a random Haar unitary one needs an exponential number of two-qubit gates and random bits [59]. Luckily, our encoding and decoding do not require sampling from the Haar measure, but rather only from a distribution which agrees with it up to the tth moment, for appropriate t. Such equivalent distributions are known as unitary t-designs [37–44], and are more efficient to implement.

Specifically, let U be a unitary representation on $\mathbb{C}^{d \times d}$ and $P_{t,t}(U)$ be a matrix whose entries are polynomials of order t in the coefficients of U and of order t in the coefficients of U^*, and let $\mathbb{E}_{U \sim \nu}[f(U)]$ be the expectation value of a function f according to measure ν. We then say that $\mathbb{E}_{U \sim \nu_{\text{Haar}}}[P_{t,t}(U)]$, where ν_{Haar} denotes the Haar measure, *admits a unitary t-design*.

Proposition 3 (Encoder and decoder as designs). The encoder $\mathcal{E}_{\text{cov}}(X_{\text{L}})$ and decoder $\mathcal{D}_{\text{cov}}(X_{\text{C}})$ admit unitary $(n_{\text{C}} + n_{\text{R}}/2 + 1)$-designs for all $X_{\text{L}} \in \text{Lin}(\mathcal{H}_{\text{L}})$, $X_{\text{C}} \in \text{Lin}(\mathcal{H}_{\text{C}}^{\otimes n})$.

Proof. We start with the encoder. Expanding the definition Eq. (3), one can write

$$\mathcal{E}_{\text{cov}}(\cdot) = \mathbb{E}_{U \sim \text{Haar}}[(U^{\otimes n_{\text{C}}} \mathcal{E}(U_{\text{L}}^{\dagger}(\cdot)U_{\text{L}})U^{\dagger \otimes n_{\text{C}}})$$
$$\otimes ((I \otimes U)^{\otimes n_{\text{R}}/2} \psi (I \otimes U^{\dagger})^{\otimes n_{\text{R}}/2})]. \quad \text{(D4)}$$

Recall from Sec. A 1 that the representation of U used on all the individual qudits is the same faithful representation of $\text{SU}(d)$. As such, the logical and computational representations are related via a linear isomorphism, $V_{\text{C} \to \text{L}}$, satisfying $U_{\text{L}} = V_{\text{C} \to \text{L}} U V_{\text{C} \to \text{L}}^{\dagger}$ for all logical and computational unitary representations U_{L} and U. Since the Hilbert spaces involved are finite dimensional, and $V_{\text{C} \to \text{L}}$, $\mathcal{E}$ are linear maps, it follows that the matrix entries of the terms in square brackets in Eq. (D4), are polynomials of order $(n_{\text{C}} + n_{\text{R}}/2 + 1)$ in the coefficients of U and of the same order in the matrix entries of U^*. Hence it is a unitary $(n_{\text{C}} + n_{\text{R}}/2 + 1)$-design.

In the case of the decoder, we have

$$\mathcal{D}_{\text{cov}}(\cdot) = \mathbb{E}_{U \sim \text{Haar}}[\mathcal{U}_{\text{L}} \circ \mathcal{D} \circ \mathcal{U}_{\text{C}}^{-1} \otimes \mathcal{M}_U](\cdot), \quad \text{(D5)}$$

where

$$\mathcal{M}_U(\cdot) := \sum_{\mathbf{s} \subset \{1,\ldots,s_{\text{R}}\}} \text{Tr}\,[(\cdot)|\eta_U\rangle\langle\eta_U|_{\mathbf{s}} \otimes P_{\text{err},\mathbf{s}^c}], \quad \text{(D6)}$$

with $|\eta_U\rangle = (U \otimes I)^m |\eta_0\rangle$, and $|\eta_0\rangle := \bigoplus_\lambda d_\lambda |\Phi_\lambda^+\rangle \otimes |\Phi_{m_\lambda}^+\rangle$, $s_{\text{R}} \cdot 2m = n_{\text{R}}$. Similarly to as in Eq. (D4), the term $\mathcal{U}_{\text{L}} \circ \mathcal{D} \circ \mathcal{U}_{\text{C}}^{-1}(\cdot)$ in Eq. (D5), when evaluated on any input, has matrix entries which are polynomials of order $n + 1$ in the coefficients of U and also of the same order in the coefficients of U^*. Hence observing the form of Eq. (D6), we conclude the proof. ∎

Reference [43] devises and quantifies a method to approximate unitary t-designs. We start with a strategy to construct a random unitary U over N qubits on sites labeled 1 to N: Pick an index l uniformly from $[N - 1]$ and a unitary denoted $U_{l,l+1}$, drawn from the Haar measure on $\text{SU}(4)$, which acts on the two neighboring qubits, l and $l + 1$. Repeat the above k times and multiply the unitaries together. The resultant unitary is U and it is sampled from a distribution which we denote $\nu(k)$.

In our case if the qudits each consist of N qubits, we can use this procedure to construct an approximate covariant encoding as follows: We record the random sequence of k nearest neighbor unitaries $U_{l,l+1}$. We first apply it to the logical qubits and encode via $\mathcal{E}$, and then apply the recorded random sequence $n_{\text{C}} + n_{\text{R}}/2$ times to the computational and reference frame qubits. The resultant encoder is

$$\mathcal{E}_{\text{cov}}^{\nu(k)}(\cdot) := \mathbb{E}_{U \sim \nu(k)}[\mathcal{U}_{\text{C}} \circ \mathcal{E} \circ \mathcal{U}_{\text{L}}^{-1}(\cdot) \otimes \mathcal{U}_{\text{R}}(\Psi)]. \quad \text{(D7)}$$

Similarly, one can use the procedure to produce the approximate decoder,

$$\mathcal{D}_{\text{cov}}^{\nu(k)}(\cdot) := \mathbb{E}_{\hat{U} \sim \nu(k)}[(\hat{\mathcal{U}}_{\text{L}} \circ \mathcal{D} \circ \hat{\mathcal{U}}_{\text{C}}^{-1} \otimes \mathcal{M}_{\hat{U}})(\cdot)]. \quad \text{(D8)}$$

From [43] it follows that

$$\left\| \mathcal{E}_{\text{cov}}^{\nu(k)} - \mathcal{E}_{\text{cov}} \right\|_\diamond \leqslant \epsilon_{\text{cov}}, \quad \text{(D9)}$$

$$\left\| \mathcal{D}_{\text{cov}}^{\nu(k)} - \mathcal{D}_{\text{cov}} \right\|_\diamond \leqslant \epsilon_{\text{cov}}, \quad \text{(D10)}$$

if

$$k = 170,000\, N \lceil \log(4(n_{\text{C}} + n_{\text{R}}/2 + 1)) \rceil^2 (n_{\text{C}} + n_{\text{R}}/2 + 1)^{8.1}$$
$$\times (2N(n_{\text{C}} + n_{\text{R}}/2 + 1) + 1 + \log(1/\epsilon_{\text{cov}})), \quad \text{(D11)}$$

which scales polynomially in both n_{C} and n_{R} (recall lower bounds on ϵ_{cov}, in Propositions 1 and 2). Since sampling a polynomial number of times from $\text{SU}(4)$ can be performed efficiently, both the approximate encoder $\mathcal{E}_{\text{cov}}^{\nu(k)}$ and decoder $\mathcal{D}_{\text{cov}}^{\nu(k)}$ can be efficiently implemented in n_{C} and n_{R} so long as the reference frame state Ψ and measurement $\mathcal{M}_{\hat{U}}$ can be efficiently constructed. It is also important to note that, under the application of an arbitrary number of transversal gates, the errors in Eqs. (D9), (D10) do not grow [44].

APPENDIX E: DETAILS OF THE NUMERICAL CALCULATIONS

In this section, we introduce the numerical experiments we implemented.

We use the "five-qubit code" [54] (5 computational qubits; one logical qubit) as the subroutine encoding and decoding pair, and calculate the performance of our $\text{SU}(2)$ covariant code. The "five-qubit code" can correct arbitrary single qubit errors on its code space. It is realized by the following encoding:

$$|0_{\text{L}}\rangle \to 1/4(|00000\rangle + |10010\rangle + |01001\rangle + |10100\rangle$$
$$+ |01010\rangle - |11011\rangle - |00110\rangle - |11000\rangle$$
$$- |11101\rangle - |00011\rangle - |11110\rangle - |01111\rangle$$
$$- |10001\rangle - |01100\rangle - |10111\rangle + |00101\rangle),$$
$$|1_{\text{L}}\rangle \to 1/4(|11111\rangle + |01101\rangle + |10110\rangle + |01011\rangle$$
$$+ |10101\rangle - |00100\rangle - |11001\rangle - |00111\rangle$$
$$- |00010\rangle - |11100\rangle - |00001\rangle - |10000\rangle$$
$$- |01110\rangle - |10011\rangle - |01000\rangle + |11010\rangle).$$

After this encoding, arbitrary single qubit errors can be detected with stabilizer measurements with 4 ancilla qubits. Depending on the measurement outcome, a correction will be performed, and then the original logical qubit will be obtained by doing the inverse of the encoding map.

In our experiment, not only erasure errors but also other common error types like dephasing errors and depolarizing errors are considered. For erasure errors, we use the generalized sine states as the reference frame state. This is given by $|\phi\rangle \propto \sum_{j=j_{\min}}^{J} \frac{\sin(\pi j/J)}{\sqrt{2j+1}} |I^{(j)}\rangle\rangle$. After encoding with U_g, it becomes $|\phi_g\rangle \propto \sum_{j=j_{\min}}^{J} \frac{\sin(\pi j/J)}{\sqrt{2j+1}} |U_g^{(j)}\rangle\rangle$. The probability density function of the outcome h with covariant POVM $\{dU_h, |\eta_h\rangle\langle\eta_h|\}$,

$|\eta_h\rangle = \oplus_j |U_h^{(j)}\rangle\rangle$ is then given by:

$$p(h|g) = \frac{\left| \sum_j \frac{\sin(\pi j/J)}{\sqrt{2j+1}} \operatorname{Tr}\left[U_{gh^{-1}}^{(j)}\right]\right|^2}{\left| \sum_j \sin(\pi j/J) \cdot \sqrt{2j+1}\right|^2}. \qquad \text{(E1)}$$

For i.i.d. depolarizing and dephasing errors, we use $n_R/2$ Bell states $|\phi\rangle = |\Phi^+\rangle^{\otimes n_R/2}$ as the reference frame state. After encoding with U_g, it becomes $|\phi_g\rangle = |\Phi_g^+\rangle^{\otimes n_R/2}$. When measuring these states, we will get a set of U_h with $n_R/2$ elements: $\{U_{h,1}, ..., U_{h,n_R/2}\}$. From this set we will get a final U_h, which maximizes $\langle\Phi_h^+|(\sum_{i=1}^{n_R/2} |\Phi_{h,i}^+\rangle\langle\Phi_{h,i}^+|)|\Phi_h^+\rangle$, and can be regarded as the best representation of these measurement outcomes.

We also tested our protocol in the situation where one fifth of the reference frame qubits go through the completely depolarizing/dephasing error channel. For this case, an algorithm similar as majority voting is first used. For each $U_{h,i}$, we calculate $\langle\Phi_{h,i}^+|(\sum_{i=1}^{n_R/2} |\Phi_{h,i}^+\rangle\langle\Phi_{h,i}^+|)|\Phi_{h,i}^+\rangle$. The smallest one fifth outcomes will be regarded as affected by noise. The remaining set will be used to get the final U_h which is the best representation of these outcomes.

After getting U_h, we perform the decoding according to it and then calculate the performance. We analyze how the performance changes as a function of the size of the reference frame state.

To calculate the performance, we first calculate the entanglement fidelity as defined in Eq. (A2). For the five-qubit code, we get:

$$F_{\text{ent},5-\text{code}} = \int dg \, p(h|g) \, F_{\text{ent},5-\text{code}}(g, h), \qquad \text{(E2)}$$

where $p(h|g)$ is the probability to get measurement outcome U_h with encoding U_g, and $F_{\text{ent},5-\text{code}}(g, h)$ is the entanglement fidelity using U_g in encoding and U_h in decoding. For every reference frame size considered ($2m$ for the sine state and n_R for the Bell states), we randomly generated 400 unitaries U_g from the Haar measure, and calculated their average entanglement fidelity, giving us the final $F_{\text{ent},5-\text{code}}$.

To calculate the worst-case error $\epsilon_{\text{cov},5-\text{code}}$, we use the relationship between the worst-case error and the entanglement error in Eq. (A7). Notice that the whole process of encoding, noisy evolution, and decoding in the considered case is a qubit covariant channel. Therefore its Choi state can always be decomposed as $\lambda|\Phi^+\rangle\langle\Phi^+| + \frac{(1-\lambda)}{3}(I - |\Phi^+\rangle\langle\Phi^+|)$, for a $\lambda \in [0, 1]$. Thus its worst-case error is equal to one minus the entanglement fidelity. Thus we can bound the worst-case error by $\epsilon_{\text{cov},5-\text{code}} \leqslant 2 \cdot (1 - F_{\text{ent},5-\text{code}})$.

---

[1] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature **549**, 172 (2017).

[2] P. W. Shor, Fault-tolerant quantum computation, in *Proceedings of the 37th Annual Symposium on Foundations of Computer Science, Burlington, VT, USA* (IEEE, Piscataway, NJ, 1996), pp. 56–65.

[3] D. Gottesman, Quantum error correction and fault tolerance, in *Encyclopedia of Mathematical Physics*, edited by J.-P. Françoise, G. L. Naber, and T. S. Tsun (Academic Press, Oxford, 2006), pp. 196–201.

[4] B. Eastin and E. Knill, Restrictions on Transversal Encoded Quantum Gate Sets, Phys. Rev. Lett. **102**, 110502 (2009).

[5] M. P. Woods and Á. M. Alhambra, Continuous groups of transversal gates for quantum error correcting codes from finite clock reference frames, Quantum **4**, 245 (2020).

[6] P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill, Continuous Symmetries and Approximate Quantum Error Correction, Phys. Rev. X **10**, 041018 (2020).

[7] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[8] E. Knill, R. Laflamme, and W. Zurek, Threshold accuracy for quantum computation, arXiv:quant-ph/9610011.

[9] H. Bombin and M. A. Martin-Delgado, Topological Computation without Braiding, Phys. Rev. Lett. **98**, 160502 (2007).

[10] A. Paetznick and B. W. Reichardt, Universal Fault-Tolerant Quantum Computation with Only Transversal Gates and Error Correction, Phys. Rev. Lett. **111**, 090505 (2013).

[11] T. Jochym-O'Connor and R. Laflamme, Using Concatenated Quantum Codes for Universal Fault-Tolerant Quantum Gates Phys. Rev. Lett. **112**, 010505 (2014).

[12] H. Bombín, Gauge color codes: Optimal transversal gates and gauge fixing in topological stabilizer codes, New J. Phys. **17**, 083002 (2015).

[13] T. J. Yoder, R. Takagi, and I. L. Chuang, Universal Fault-Tolerant Gates on Concatenated Stabilizer Codes, Phys. Rev. X **6**, 031039 (2016).

[14] B. J. Brown, A fault-tolerant non-Clifford gate for the surface code in two dimensions, Sci. Adv. **6**, eaay4929 (2020).

[15] A. Peres and P. F. Scudo, Transmission of a Cartesian Frame by a Quantum System, Phys. Rev. Lett. **87**, 167901 (2001).

[16] E. Bagan, M. Baig, and R. Muñoz-Tapia, Aligning Reference Frames with Quantum States, Phys. Rev. Lett. **87**, 257903 (2001).

[17] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Efficient Use of Quantum Resources for the Transmission of a Reference Frame, Phys. Rev. Lett. **93**, 180503 (2004).

[18] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Reference frames, superselection rules, and quantum information, Rev. Mod. Phys. **79**, 555 (2007).

[19] A. Kitaev, D. Mayers, and J. Preskill, Superselection rules and quantum protocols, Phys. Rev. A **69**, 052326 (2004).

[20] I. Marvian and R. W. Spekkens, The theory of manipulations of pure state asymmetry: I. Basic tools, equivalence classes and single copy transformations, New J. Phys. **15**, 033001 (2013).

[21] I. Marvian and R. B. Mann, Building all time evolutions with rotationally invariant Hamiltonians, Phys. Rev. A **78**, 022304 (2008).

[22] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Dialogue concerning two views on quantum coherence: factist and fictionist, Int. J. Quantum. Inform. **04**, 17 (2006).

[23] R. M. Angelo, N. Brunner, S. Popescu, A. J. Short, and P. Skrzypczyk, Physics within a quantum reference frame, J. Phys. A: Math. Theor. **44**, 145304 (2011).

[24] R. M. Angelo and A. D. Ribeiro, Kinematics and dynamics in noninertial quantum frames of reference, J. Phys. A: Math. Theor. **45**, 465306 (2012).

[25] M. Krumm, P. A. Höhn, and M. P. Müller, Quantum reference frame transformations as symmetries and the paradox of the third particle, Quantum **5**, 530 (2021).

[26] P. A. Höhn, A. R. H. Smith, and M. P. E. Lock, The trinity of relational quantum dynamics, Phys. Rev. D **104**, 066001 (2021).

[27] F. Giacomini, E. Castro-Ruiz, and Č. Brukner, Quantum mechanics and the covariance of physical laws in quantum reference frames, Nat. Commun. **10**, 494 (2019).

[28] A. Vanrietvelde, P. A. Höhn, and F. Giacomini, Switching quantum reference frames in the *N*-body problem and the absence of global relational perspectives, arXiv:1809.05093.

[29] F. Giacomini, E. Castro-Ruiz, and Č. Brukner, Relativistic Quantum Reference Frames: The Operational Meaning of Spin, Phys. Rev. Lett. **123**, 090404 (2019).

[30] A. Vanrietvelde, P. A. Hoehn, F. Giacomini, and E. Castro-Ruiz, A change of perspective: switching quantum reference frames via a perspective-neutral framework, Quantum **4**, 225 (2020).

[31] A.-C. de la Hamette, T. D. Galley, P. A. Hoehn, L. Loveridge, and M. P. Mueller, Perspective-neutral approach to quantum frame covariance for general symmetry groups, arXiv:2110.13824.

[32] P. A. Hoehn, M. P. E. Lock, S. A. Ahmad, A. R. H. Smith, and T. D. Galley, Quantum relativity of subsystems, arXiv:2103.01232.

[33] A.-C. de la Hamette and T. D. Galley, Quantum reference frames for general symmetry groups, Quantum **4**, 367 (2020).

[34] P. Hayden, S. Nezami, S. Popescu, and G. Salton, Error Correction of Quantum Reference Frame Information, PRX Quantum **2**, 010326 (2021).

[35] A. Acín, E. Jané, and G. Vidal, Optimal estimation of quantum dynamics, Phys. Rev. A **64**, 050302(R) (2001).

[36] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Optimal estimation of group transformations using entanglement, Phys. Rev. A **72**, 042338 (2005).

[37] R. Oliveira, O. C. O. Dahlsten, and M. B. Plenio, Generic Entanglement Can Be Generated Efficiently, Phys. Rev. Lett. **98**, 130502 (2007).

[38] O. C. O. Dahlsten, R. Oliveira, and M. B. Plenio, The emergence of typical entanglement in two-party random processes, J. Phys. A: Math. Theor. **40**, 8081 (2007).

[39] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[40] A. W. Harrow and R. A. Low, Random Quantum Circuits are Approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[41] I. T. Diniz and D. Jonathan, Comment on "Random quantum circuits are approximate 2-designs" by A. W. Harrow and R. A. Low (Commun. Math. Phys. 291, 257 (2009)), Commun. Math. Phys. **304**, 281 (2011).

[42] L. Arnaud and D. Braun, Efficiency of producing random unitary matrices with quantum circuits, Phys. Rev. A **78**, 062329 (2008).

[43] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[44] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, Efficient Quantum Pseudorandomness with Nearly Time-Independent Hamiltonian Dynamics, Phys. Rev. X **7**, 021006 (2017).

[45] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, El Paso Texas USA*, STOC '97 (ACM Press, New York, 1997), pp. 176–188.

[46] V. Bužek, R. Derka, and S. Massar, Optimal Quantum Clocks, Phys. Rev. Lett. **82**, 2207 (1999).

[47] A. Kubica and R. Demkowicz-Dobrzański, Using Quantum Metrological Bounds in Quantum Error Correction: A Simple Proof of the Approximate Eastin-Knill Theorem, Phys. Rev. Lett. **126**, 150503 (2021).

[48] S. Zhou, Z.-W. Liu, and L. Jiang, New perspectives on covariant quantum error correction, Quantum **5**, 521 (2021).

[49] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Capacities of Quantum Erasure Channels, Phys. Rev. Lett. **78**, 3217 (1997).

[50] D. Gottesman, Stabilizer Codes and Quantum Error Correction, PhD. thesis, California Institute of Technology, May 1997.

[51] J. M. Renes, F. Dupuis, and R. Renner, Efficient Polar Coding of Quantum Information, Phys. Rev. Lett. **109**, 050504 (2012).

[52] J. M. Renes and M. M. Wilde, Polar Codes for Private and Quantum Communication Over Arbitrary Channels, IEEE Trans. Inf. Theory **60**, 3090 (2014).

[53] S. Kumar, R. Calderbank, and H. D. Pfister, Reed-muller codes achieve capacity on the quantum erasure channel, in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain* (IEEE, Piscataway, NJ, 2016), pp. 1750–1754.

[54] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Perfect Quantum Error Correcting Code, Phys. Rev. Lett. **77**, 198 (1996).

[55] Y. Mo, COVARIANTQEC: Code for optimal universal covariant quantum error correction (2021), https://github.com/moyin23/covariantqec.

[56] Y. Yang, G. Chiribella, and D. Ebler, Efficient Quantum Compression for Ensembles of Identically Prepared Mixed States, Phys. Rev. Lett. **116**, 080501 (2016).

[57] Y. Yang, G. Chiribella, and M. Hayashi, Optimal Compression for Identically Prepared Qubit States, Phys. Rev. Lett. **117**, 090502 (2016).

[58] Y. Yang, G. Bai, G. Chiribella, and M. Hayashi, Compression for quantum population coding, IEEE Trans. Inf. Theory **64**, 4766 (2018).

[59] E. Knill, Approximation by quantum circuits, arXiv:quant-ph/9508006.

[60] M. Oszmaniec, A. Sawicki, and M. Horodecki, Epsilon-nets, unitary designs and random quantum circuits, IEEE Trans. Inf. Theory **68**, 989 (2022).

[61] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[62] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[63] D.-S. Wang, G. Zhu, C. Okay, and R. Laflamme, Quasi-exact quantum computation, Phys. Rev. Research **2**, 033116 (2020).

[64] F. Albarelli, M. Barbieri, M. G. Genoni, and I. Gianani, A perspective on multiparameter quantum metrology: From

theoretical tools to applications in quantum imaging, Phys. Lett. A **384**, 126311 (2020).

[65] W. Gorecki, S. Zhou, L. Jiang, and R. Demkowicz-Dobrzański, Optimal probes and error-correction schemes in multi-parameter quantum metrology, Quantum **4**, 288 (2020).

[66] Z. Hou, Z. Zhang, G.-Y. Xiang, C.-F. Li, G.-C. Guo, H. Chen, L. Liu, and H. Yuan, Minimal Tradeoff and Ultimate Precision Limit of Multiparameter Quantum Magnetometry under the Parallel Scheme, Phys. Rev. Lett. **125**, 020501 (2020).

[67] A. Z. Goldberg, L. L. Sánchez-Soto, and H. Ferretti, Intrinsic Sensitivity Limits for Multiparameter Quantum Metrology, Phys. Rev. Lett. **127**, 110501 (2021).

[68] Johannes J. Meyer, J. Borregaard, and J. Eisert, A variational toolbox for quantum multi-parameter estimation, npj Quantum Inf. **7**, 1 (2021).

[69] Y. Yang, R. Renner, and G. Chiribella, Optimal Universal Programming of Unitary Gates, Phys. Rev. Lett. **125**, 210501 (2020).

[70] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, J. High Energy Phys. 04 (2015) 163.

[71] D. Harlow and H. Ooguri, Constraints on Symmetries from Holography, Phys. Rev. Lett. **122**, 191601 (2019).

[72] D. Harlow and H. Ooguri, Symmetries in quantum field theory and quantum gravity, Commun. Math. Phys. **383**, 1669 (2021).

[73] T. Kohler and T. Cubitt, Toy models of holographic duality between local hamiltonians, J. High Energy Phys. 08 (2019) 017.

[74] A. Milekhin, Quantum error correction and large N, SciPost Phys. **11**, 094 (2021).

[75] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, Rev. Mod. Phys. **84**, 621 (2012).

[76] W. Fulton and J. Harris, *Representation Theory: A First Course*, *volume 129 of Graduate Texts in Mathematics* (Springer, New York, 2004).

[77] M. Raginsky, A fidelity measure for quantum channels, Phys. Lett. A **290**, 11 (2001).

[78] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surveys **52**, 1191 (1997).

[79] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).

[80] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[81] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, UK, 2018).

[82] D. Bacon, I. L. Chuang, and A. W. Harrow, Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms, Phys. Rev. Lett. **97**, 170502 (2006).

[83] H. Krovi, An efficient high dimensional quantum schur transform, Quantum **3**, 122 (2019).

[84] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2010).

[85] J. Watrous, Simpler semidefinite programs for completely bounded norms, Chicago J. Theor. Comput. Sci. **2013**, 8 (2013).

[86] N. Killoran, Entanglement Quantification and Quantum Benchmarking of Optical Communication Devices, PhD. thesis, University of Waterloo, 2012.

[87] L. P. Thinh, P. Faist, J. Helsen, D. Elkouss, and S. Wehner, Practical and reliable error bars for quantum process tomography, Phys. Rev. A **99**, 052311 (2019).

[88] V. Katariya and M. M. Wilde, Geometric distinguishability measures limit quantum channel estimation and discrimination, Quantum Inf. Proc. **20**, 1 (2021).

[89] R. Alicki, S. Rudnicki, and S. Sadowski, Symmetry properties of product states for the system of Nn-level atoms, J. Math. Phys. **29**, 1158 (1988).

[90] M. Christandl and G. Mitchison, The spectra of quantum states and the Kronecker coefficients of the symmetric group, Commun. Math. Phys. **261**, 789 (2006).

[91] R. O'Donnell and J. Wright, Quantum spectrum testing, in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing (STOC'15)* (ACM Press, New York, NY, 2015), pp. 529–538.

[92] M. Hayashi, *A Group Theoretic Approach to Quantum Information* (Springer International Publishing, Cham, 2017).