

Achievable Rates for Concatenated Square Gottesman-Kitaev-Preskill Codes

Mahadevan Subramanian^{✉,*}, Guo Zheng^{✉,†} and Liang Jiang^{✉,‡}

Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA



(Received 28 May 2025; revised 3 September 2025; accepted 20 October 2025; published 21 November 2025)

The Gottesman-Kitaev-Preskill (GKP) codes are known to achieve optimal rates under displacement noise and pure-loss channels, which establishes theoretical foundations for its optimality. However, such optimal rates are only known to be achieved at a discrete set of noise strengths with the current self-dual symplectic lattice construction. In this work, we develop a new coding strategy using concatenated continuous variable-discrete variable encodings to go beyond past results and establish GKP's optimal rate over all noise strengths. In particular, for displacement noise, the rate is obtained through a constructive approach by concatenating GKP codes with a quantum polar code and analog decoding. For a pure-loss channel, we prove the existence of capacity-achieving GKP codes through a random coding approach. These results highlight the capability of concatenation-based GKP codes and provides new methods for constructing good GKP lattices.

DOI: [10.1103/56vj-z7h1](https://doi.org/10.1103/56vj-z7h1)

I. INTRODUCTION

The goal of reliably communicating quantum information through noisy channels lies at the heart of proposals for quantum networks [1–5] as well as quantum transduction [6]. A key aspect of this is understanding what rates of information transmission can be considered achievable [7], which has laid the foundation for the study of the quantum capacities of noisy channels [8,9]. Showing a rate to be achievable offers a direct lower bound to the capacity of the noise channel. However, explicitly showing the achievability of rates requires proving the existence of a sequence of encoding and decoding schemes that are asymptotically reliable for that particular rate. Due to the non-trivial nature of this task, the coherent information of a channel is often used to find achievable rates. The coherent information is the quantum analog to mutual information and offers a lower bound to the quantum capacity [9]. For degradable channels such as the bosonic pure-loss and amplification channels, the maximal one-shot coherent information equals the capacity. Upper and lower bounds on the capacity for channels such as the Gaussian displacement noise [10,11], loss-dephasing channel [12] and

thermal loss channels [13] have been found while the exact capacity remains unknown.

Gottesman-Kitaev-Preskill (GKP) codes [14] are bosonic error correcting codes that encode a finite-dimensional Hilbert space into bosonic modes. These codes are obtained by using a stabilizer group generated by optical phase space displacements, and hence are deeply connected to symplectically integral lattices [15,16]. This connection has been used to show that GKP codes can achieve the coherent information of the Gaussian displacement noise channel [11] as well as the capacity of the pure-loss and amplification channels [17]. However, these results have two major limitations. The rate R (in qubit per mode) these codes achieve are restricted to have 2^R be an integer, and these are existence-based results which do not provide constructive ways to obtain rate achieving codes. As a result, the optimal rates are only achieved for a discrete set of noise strengths. All known asymptotic GKP coding rate results [11,17] make use of the family of lattices obtained by rescaling self-dual (unimodular) lattices which as a corollary of the Buser-Sarnak theorem [18], will always contain a good spherical packing lattice. This good spherical packing can be used to prove achievability for certain rates. However, to encode information self-dual lattices must be scaled by a square root of an integer (ensuring that it remains valid for encoding a GKP code), which forces the number of logical dimensions per mode to be an integer. Searching for the lattices that satisfy these properties is a hard task [19] and the current knowledge of optimal sphere-packing lattices is limited to only dimensions of 24 [20]. While the capacity achieving lattices need not be truly optimal in regard to sphere-packing, this does

*Contact author: mahadevans@uchicago.edu

†Contact author: guozheng@uchicago.edu

‡Contact author: liangjiang@uchicago.edu

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

highlight the difficulty one faces in trying to construct such lattices.

In this work, we develop a different approach to overcome this limitation, enabling the achievement of favorable rates through more explicit construction methods. We focus on GKP codes that are obtained by taking N single mode square GKP codes, each encoding a qudit of d levels and then defining an outer code for concatenating as a N qudit stabilizer code encoding K qudits which we denote by $[[N, K]]_d$. Phase space displacements representing logical Pauli operations for the single mode square GKP code are spaced apart by $\sqrt{2\pi/d}$. Hence, the effective $2N$ -dimensional lattice describing the multimode code obtained through this concatenation is choosing from an increasingly finer grid of displacements for the stabilizer generators as d increases. Heuristically, this can be used to mimic behaviors of good spherical-packing lattices while also achieving optimal rates over all noise strengths since these concatenated codes are no longer restricted to having 2^R be an integer (where R is the achievable rate). This motivates our study for concatenation to GKP qudits of increasing dimension d . For the Gaussian displacement noise channel, we first find that the effective noise channel for a single-mode GKP square qudit is a classical mixture of Pauli noise where knowledge of the syndrome gives us knowledge of what the Pauli noise looks like. We formalize this by finding an achievable rate for this noise channel with the use of quantum polar codes [21] that take advantage of the analog information offered by this syndrome. The usefulness of GKP analog information has been well studied [22,23], to which we show a rigorous way to quantify the advantage analog information offers. We find that this newly defined achievable rate approaches the coherent information of the Gaussian displacement noise channel at all values of noise strength σ . Crucially, these GKP codes can be explicitly constructed, and also have efficient ($\mathcal{O}(N \log N)$ circuit depth) encoding and decoding procedures due to the outer concatenation with the quantum polar code. We further support this with a numerical study of the performance of these polar codes for increasing numbers of modes under the Gaussian displacement noise model. Different from the existing work for constructing constant rate N mode GKP codes with shortest non-zero lattice vector scaling of $\Omega(\sqrt{N})$ [24], our work explicitly constructs lattices that can achieve a rate approaching the coherent information for displacement noise channel.

We also show that concatenated GKP codes achieve the capacity of pure-loss, and hence by extension also the capacity of the amplification channel [17]. We use a construction of stabilizer codes for prime d -dimensional qudits mapped from self-orthogonal codes in the Galois field $GF(d^2)$ from Ref. [25]. By averaging over the set of GKP codes obtained in this manner, we prove the existence of a capacity achieving sequence of codes as $d \rightarrow \infty$, in essence recreating the behavior of a

good spherical-packing lattice. These results highlight the importance of the family of codes obtained by concatenating qudit codes to square GKP codes, as well as their capability in exceeding previously known achievable rates for GKP codes.

This paper is structured as follows. In Sec. II A, we introduce the square GKP code, following which we provide a short introduction to achievable rates for communication of quantum information in Sec. II B. Our results begin with Sec. III A where we show how the effective logical noise channel of a corrected square GKP qudit is a classical mixture of Pauli noise with an auxiliary output containing syndrome information. We then show in Sec. III B how the coherent information of this logical noise channel which we refer to as $I_{d,\text{analog}}^{\text{sq}}$ approaches the coherent information of the Gaussian displacement noise channel for all values of noise strength σ for large enough d . In Sec. IV A we provide a brief introduction to polar codes and highlight the construction of quantum polar codes we make use of in Sec. IV B and provide our numeric study of these codes in the context of achieving $I_{d,\text{analog}}^{\text{sq}}$ in Sec. IV C. In Sec. V A we discuss the capacity of the bosonic pure-loss and amplification channels following which we show the existence of a capacity achieving sequence for the pure-loss channel by concatenation with square GKP codes in Sec. V B. Finally we discuss the relevance and implications of our results in Sec. VI.

II. BACKGROUND AND NOTATION

A. The square GKP code

We begin our discussion focusing on the square GKP code in one mode. We make use of canonical position and momentum operators $\hat{q}$ and $\hat{p}$ respectively which satisfy the commutation relation $[\hat{q}, \hat{p}] = i$. The stabilizers are chosen as phase space displacements of equal length in the direction of $\hat{q}$ and $\hat{p}$. We consider the two stabilizers

$$\hat{S}_1 = \exp(i\hat{q}\sqrt{2\pi d}), \quad \hat{S}_2 = \exp(-i\hat{p}\sqrt{2\pi d}), \quad (1)$$

where d is some prime number. For this stabilizer set, the displacements which commute with both $\hat{S}_1$ and $\hat{S}_2$ can be written as

$$\hat{P}_{u,v} = \exp\left(i(-u\hat{p} + v\hat{q})\sqrt{\frac{2\pi}{d}}\right), \quad (2)$$

where $u, v \in \mathbb{Z}$. This then gives logical operators $\hat{X}_L = \hat{P}_{1,0}$ and $\hat{Z}_L = \hat{P}_{0,1}$ which can be used to construct the Pauli group. To see how, we can note that $\hat{X}_L \hat{Z}_L = e^{-2\pi i/d} \hat{Z}_L \hat{X}_L$ giving the necessary commutation relation for the qudit Pauli group [26]. A phase space representation of the points depicting the displacements that make up the stabilizer group (black dots) and the set of logical operators

(pink dots) is shown in Fig. 1(a). The eigenstates of $\hat{Z}_L$ and $\hat{X}_L$ would then be

$$\begin{aligned} |j_L\rangle &= \sum_{n=-\infty}^{\infty} \left| q = (dn + j)\sqrt{2\pi/d} \right\rangle, \\ |\tilde{j}_L\rangle &= \sum_{n=-\infty}^{\infty} \left| p = (dn + j)\sqrt{2\pi/d} \right\rangle, \end{aligned} \quad (3)$$

which would satisfy

$$\hat{X}^u |j_L\rangle = |j \oplus u\rangle_L, \quad \hat{Z}^v |j_L\rangle = e^{\frac{2\pi i}{d} v j} |j_L\rangle. \quad (4)$$

To construct the Pauli group out of $\hat{P}_{u,v}$, we restrict $u, v \in \mathbb{F}_d$ which is the finite-field associated to the prime number d . We will be restricting our analysis to prime d since it allows us to ensure that we can define the Galois-field $GF(d)$ [27]. Hence $\hat{P}_{u,v}$ can be understood as the logical operation $\hat{X}^u \hat{Z}^v$ (up to a phase) and hence generates the qudit Pauli group $\mathcal{P}_d$. Let the state $|\psi\rangle$ be stabilized by both $\hat{S}_1$ and $\hat{S}_2$. We now consider an erroneous displacement to act on this state given by

$$\hat{E} = \exp \left(i(-e_1 \hat{p} + e_2 \hat{q}) \sqrt{\frac{2\pi}{d}} \right), \quad (5)$$

giving the displaced state $|\psi_e\rangle = \hat{E} |\psi\rangle$. It then follows that

$$\hat{S}_1 |\psi_e\rangle = \exp(2\pi i e_1) |\psi_e\rangle, \quad \hat{S}_2 |\psi_e\rangle = \exp(-2\pi i e_2) |\psi_e\rangle, \quad (6)$$

which means that measuring the eigenvalues of the stabilizer is equivalent to measuring the two values

$$s_1 = \left(\hat{q} \sqrt{\frac{d}{2\pi}} \right) \bmod 1, \quad s_2 = \left(-\hat{p} \sqrt{\frac{d}{2\pi}} \right) \bmod 1, \quad (7)$$

since the eigenvalues lie on a rotor for displacement operations. Importantly, each pair of values s_1, s_2 define an orthogonal subspace of the full Hilbert space (also referred to as the Zak basis [28,29]). There are broadly two ways of measuring the stabilizers for correction, namely the Steane-type measurement and the Knill (teleportation) scheme. The first is the Steane-inspired scheme introduced in Ref. [30] which makes use of CSUM gates and one GKP ancilla with homodyne measurement. The second is the teleportation-based scheme which uses an encoded Bell pair [31,32] and a beam-splitter and squeezing operations. The teleportation-based scheme propagates errors with a smaller pre-factor compared to the Steane-type scheme which leads to better performance assuming equally noisy ancillas for both schemes [33,34].

The code states defined in Eq. (3) are infinite energy states since they are a superposition of infinitely squeezed states. To physically construct GKP states, an envelope is applied on these infinite energy states to approximate them. A commonly used envelope is the Gaussian envelope $\exp(-\Delta^2 \hat{n})$ (which can also be extended to multimode envelopes [15]) which on application to the ideal GKP codewords $|\psi_{\text{ideal}}\rangle$ would yield a finite energy state. This state can be written as a superposition of finitely squeezed states [35]. The envelope operation can be decomposed in the over-complete basis of displacement operations [14,36] as a mixture of coherent displacement errors on the GKP states that can be treated as incoherent displacement noise of spread $\sigma = \tanh(\Delta^2/2)$ by a twirling argument involving the repeated application of stabilizers [33,37,38]. Additionally, the bosonic pure-loss channel as well as thermal-loss channel can be converted to a Gaussian displacement noise channel with the appropriate application of a quantum limited amplification channel [13]. Hence, the displacement error noise model proves to be a very useful noise model to study.

B. Achievable rates for communication of quantum information

A rate R is said to be achievable for a noise channel $\mathcal{N}$ if there exists a sequence of encoding and decoding operations $\mathcal{E}_n, \mathcal{D}_n$ such that over this sequence

$$\lim_{n \rightarrow \infty} \frac{\log_2(d_L(\mathcal{E}_n, \mathcal{D}_n))}{n} = R, \quad \lim_{n \rightarrow \infty} F_e(\mathcal{D}_n \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}_n) = 1, \quad (8)$$

where d_L is the total logical dimension of the encoding and decoding operations $\mathcal{E}_n, \mathcal{D}_n$ and F_e is the channel fidelity defined as

$$F_e(\mathcal{Q}) = \langle \Phi | (\mathcal{Q} \otimes \mathcal{I}_A) (|\Phi\rangle\langle\Phi|) | \Phi \rangle, \quad (9)$$

where $|\Phi\rangle$ is the purification of the maximally mixed state and $\mathcal{I}_A$ is identity on the ancillary system. This is a metric of how much entanglement survives through the channel and is also referred to as the entanglement fidelity [39] and has strong connections to average channel fidelity as well [40,41]. This essentially tells us that asymptotically, one can reliably send R qubits of information per channel. The quantum capacity then follows naturally as the quantity C_Q such that $R \leq C_Q$ if and only if R is an achievable rate [8].

In this section, we will mainly be concerned with the coherent information of a channel which is an achievable rate that lower bounds the capacity of a channel [9]. For a quantum channel $\mathcal{N}$, we can define the complementary channel $\mathcal{N}^c$ which describes the action of the channel on the environment if the channel is modeled as a unitary interaction with an environment. Defining the Stinespring

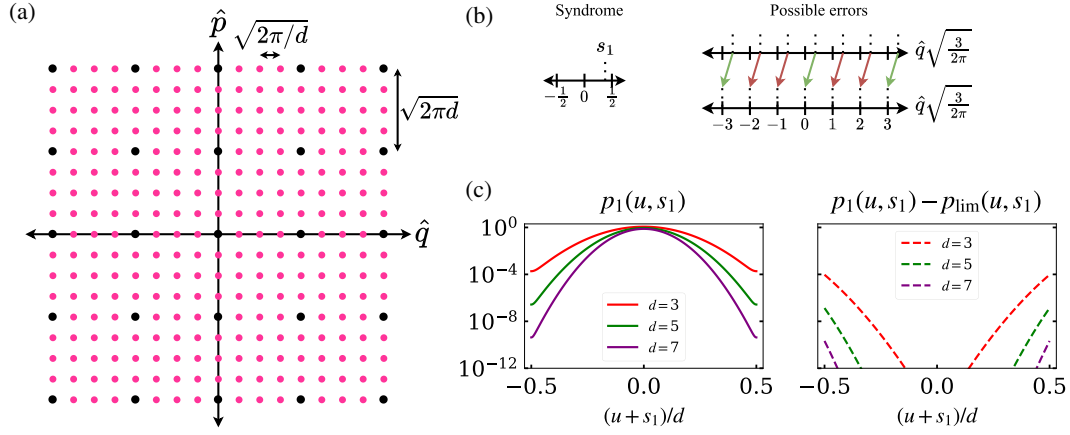


FIG. 1. (a) Depiction of the GKP lattice for a square GKP qudit of d levels. Displacements in the stabilizer group lie in the square lattice spaced by $\sqrt{2\pi d}$ (black dots) and the displacements giving logical operations are in the square lattice spaced apart by $\sqrt{2\pi/d}$ (pink dots). (b) Figure depicting the relation between the obtained syndrome and logical errors for the case of $d = 3$. Obtaining a particular syndrome s_1 implies a possible set of errors (see top right line graph) for the shift along $\hat{q}\sqrt{3/2\pi}$. The syndrome s_1 uniquely defines the smallest (in magnitude) displacement which will return the state back into the subspace with $s_1 = 0$. After this displacement has been applied, it may result in a logical error (red arrows) or result in no logical error at all (green arrows). (c) By noting that the value of $p_1(u, s_1)$ is only a function of $u + s_1$, we numerically compare these distributions for different values of d for the value of $\sigma = 0.5$. In our convention, $(u + s_1)/d$ will always lie in $[-1/2, 1/2]$ and we can see that over this normalized range, there is a strong concentration near 0 for $p_1(u, s_1)$ as d is increased. Further, we plot the difference $p_1(u, s_1) - p_{\text{lim}}(u, s_1)$ [which is also purely a function of $(u + s_1)/d$] and note that increasing the value of d shows a clear exponential suppression in this difference.

dilation of $\mathcal{N}$ as

$$\mathcal{N}(\rho_S) = \text{Tr}_E(U_{SE}(\rho_S \otimes |0\rangle\langle 0|_E)U_{SE}^\dagger), \quad (10)$$

where U_{SE} is a unitary interaction of the system and environment, we get

$$\mathcal{N}^c(\rho_S) = \text{Tr}_S(U_{SE}(\rho_S \otimes |0\rangle\langle 0|_E)U_{SE}^\dagger). \quad (11)$$

The coherent information of a channel $\mathcal{N}$ is then defined as

$$I_c(\mathcal{N}) = \sup_{\rho} (I_c(\rho, \mathcal{N})),$$

$$I_c(\rho, \mathcal{N}) = S(\mathcal{N}(\rho)) - S(\mathcal{N}^c(\rho)), \quad (12)$$

where $I_c(\rho, \mathcal{N})$ can be defined for a specific choice of input state ρ . The coherent information quantifies the maximal amount of entropy that survives through a single use of the channel [9]. For reliable information transmission through a channel, it is required that the environment does not measure out any of the information we wish to communicate. Hence through a single channel use, only as much entropy as I_c can be sent through the channel. This then gives the definition for quantum channel capacity as

$$C_Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{I_c(\mathcal{N}^{\otimes n})}{n} \geq I_c(\mathcal{N}), \quad (13)$$

which due to the additivity of $S(\rho)$ gives the lower bound.

We can link the entropic way of looking at information to an encoding operation by considering the following example. A has N Bell pairs that they wish to share the halves of with B through a noisy communication channel. If A measures an appropriate number of their halves of the Bell pair, they effectively create a stabilizer code for the halves of B with the correct choice of unitary operations. Knowing that some amount of information is inevitably gained by the environment, the measurements by A can be chosen in a way to ensure that the environment only gains redundant information that has already been measured by A . This can then allow the reliable communication (or equivalently the one-way distillation [42]) of $K \leq N$ Bell pairs effectively defining some error correcting code of form $[[N, K]]$.

III. ANALOG INFORMATION IN CONCATENATED GKP CODES

A. Effective noise channel with analog information

In this section we provide details on the effective noise channel that acts on a GKP code which has been corrected using the closest lattice point decoding for the particular syndrome output. We will mainly concern ourselves with the case of a square GKP qudit in this section and refer the reader to Appendix D to see how this extends to general GKP codes. We consider a single mode Gaussian displacement noise which can be described by the quantum

channel

$$\mathcal{N}_{\text{displ}}(\cdot) = \frac{1}{\pi\sigma^2} \int d^2\alpha e^{-|\alpha|^2/\sigma^2} D(\alpha)(\cdot)D(\alpha)^\dagger, \quad (14)$$

where $D(\alpha) = e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}$. Since this is simply a two-dimensional Gaussian random variable, we can equivalently treat this as independent Gaussian noise in the $\hat{q}$ and $\hat{p}$ quadratures respectively. Errors along $\hat{q}$ contribute to logical bit-flip and errors along $\hat{p}$ contribute to a logical phase-flip. Importantly, for a square GKP code, these are independent errors.

Consider that we now perform the stabilizer measurement associated to measuring $\hat{q}\sqrt{d/2\pi} \bmod 1$ [as per Eq. (7)] which we denote by s_1 . For our convention, we choose $x \bmod y$ to equal the smallest number (by magnitude) congruent to x which gives

$$x \bmod y = \begin{cases} x - y\lfloor x/y \rfloor & \text{if } x/y - \lfloor x/y \rfloor < 1/2 \\ x - y(1 + \lfloor x/y \rfloor) & \text{if } x/y - \lfloor x/y \rfloor \geq 1/2 \end{cases}, \quad (15)$$

which results in $s_1 \in [-1/2, 1/2)$. The smallest (in magnitude) displacement which results in the measurement result of s_1 is a shift of $s_1\sqrt{2\pi/d}$ along the $\hat{q}$ quadrature. Assuming we correct by exactly this much along the $\hat{q}$ direction, we are now left with an effective shift of $z\sqrt{2\pi/d}$ where $z \in \mathbb{Z}$. Defining $u = z \bmod d$, the shift of $z\sqrt{2\pi/d}$ along $\hat{q}$ is equivalent to a logical $\hat{X}^u$. We restrict our discussion to prime values of d . Trivially we choose the convention of $u \in \{0, 1\}$ for $d = 2$, and for $d > 2$ we define $u = z \bmod d$ similar to the modulo operation used for s_1 which gives us u to be an integer in the range $-(d-1)/2 \leq u \leq (d-1)/2$ (since d is odd). We will be frequently using equality signs involving a $\bmod d$ operation where this should be understood as being related by congruence. We will also be denoting elements of the finite field $\mathbb{F}_d$ by an integer which could for example represent a dit value. This integer should be understood equivalent to an integer in the range $[-(d-1)/2, (d-1)/2]$ to which it is congruent mod d .

We denote the probability of the event of logical $\hat{X}^u$ occurring after syndrome measurement s_1 as $p_1(u, s_1)$. To evaluate this probability, we need to find the set of displacements that cause a shift of $\hat{q} \rightarrow \hat{q} + e_1\sqrt{2\pi/d}$ such that $e_1 \bmod 1 = s_1$ (hence giving a particular syndrome) and $(e_1 - s_1) \bmod d = u$ (hence giving the logical error $\hat{X}^u$). This restricts the values of e_1 to be

$$e_1 \in \{u + s_1 + dl | l \in \mathbb{Z}\} \quad (16)$$

while the overall set of displacements resulting in this syndrome would be given by $e_1 \in \{s_1 + k | k \in \mathbb{Z}\}$. A figurative representation of the different set of displacements that

result in the same syndrome is shown in Fig. 1(b) for the case of $d = 3$. Using the knowledge of the noise model, we know that $e_1\sqrt{2\pi/d} \sim \mathcal{G}(0, \sigma^2)$ [where $\mathcal{G}(\mu, \sigma^2)$ is a Gaussian distribution of mean μ and variance σ^2] which gives

$$p(u|s_1) = \frac{\sum_{l \in \mathbb{Z}} \exp\left(-\frac{\pi}{d\sigma^2}(dl + u + s_1)^2\right)}{\sum_{k \in \mathbb{Z}} \exp\left(-\frac{\pi}{d\sigma^2}(k + s_1)^2\right)}. \quad (17)$$

Further, we can exactly find the probability distribution for the syndrome $p(s_1)$ to be

$$p(s_1) = \frac{1}{\sigma\sqrt{d}} \sum_{k \in \mathbb{Z}} \exp\left(-\frac{\pi}{d\sigma^2}(k + s_1)^2\right), \quad (18)$$

which lets us define

$$\begin{aligned} p_1(u, s_1) &= p(u|s_1)p(s_1) \\ &= \frac{1}{\sigma\sqrt{d}} \sum_{l \in \mathbb{Z}} \exp\left(-\frac{\pi d}{\sigma^2} \left(l + \frac{u + s_1}{d}\right)^2\right), \end{aligned} \quad (19)$$

showing that $p_1(u, s_1)$ is purely a function of $(u + s_1)/d$. Since u is an integer with $-(d-1)/2 \leq u \leq (d-1)/2$ (assuming d is odd and prime) and s_1 lies in the range $[-1/2, 1/2)$, the variable $(u + s_1)/d$ will satisfy $-\frac{1}{2} \leq (u + s_1)/d < \frac{1}{2}$.

The restriction of $(u + s_1)/d$ to this range by the appropriate definition of our modulo functions gives a clear understanding of what the summation over l depicts in the expression of $p_1(u, s_1)$ in Eq. (19). The terms of $l \neq 0$ are from the erroneous displacement being large enough that it displaces it outside the unit cell of the stabilizer lattice which is spaced by $\sqrt{2\pi d}$. Since the GKP code fundamentally has degenerate error sets, these can offer significant contributions depending on the form of the lattice [16]. However, in this case there is a very interesting behavior for a large value of d . The expression $p_1(u, s_1)$ shows a clear convergence towards the central term in the summation ($l = 0$) which we will denote by

$$p_{\text{lim}}(u, s_1) = \frac{1}{\sigma\sqrt{d}} \exp\left(-\frac{\pi d}{\sigma^2} \left(\frac{u + s_1}{d}\right)^2\right), \quad (20)$$

which is not a normalized distribution, but can be shown to be very close to the actual distribution for large enough d . To exactly quantify this, we show that on integrating $p_{\text{lim}}(u, s_1)$ over the variables u and s_1 (detailed calculation

in Appendix D 1), we get

$$\sum_{|u| \leq \frac{d-1}{2}} \int ds_1 p_{\text{lim}}(u, s_1) = 1 - \mathcal{O}\left(\frac{\sigma}{\sqrt{d}} e^{-\frac{d\pi}{4\sigma^2}}\right), \quad (21)$$

which gets very close to 1 as $d/\sigma^2 \gg 1$. This means that the probability contribution from all the terms in the summation of $l \neq 0$ are exponentially suppressed. Additionally increasing d makes p_{lim} have smaller spread around $(u + s_1)/d = 0$, since it is proportional to a Gaussian in the variable $(u + s_1)/d$ with spread of $\sigma\sqrt{2\pi/d}$. This can be directly observed by evaluating the full function $p_1(u, s_1)$ numerically as is plotted in Fig. 1(c).

We can similarly describe the probability distribution which relates the probability of a phase-flip error $\hat{Z}^v$ to the outcome of obtaining a syndrome measurement for s_2 denoted by $p_2(v, s_2)$. Similarly, if we consider an erroneous shift of e_2 such that $\hat{p} \rightarrow \hat{p} + e_2\sqrt{2\pi/d}$, we now have $e_2 = dl + v - s_2$ for $l \in \mathbb{Z}$. Since the random variable for e_2 follows the same distribution as e_1 , it is easy to see that the distribution

$$p_2(v, s_2) = p(u = v, s_1 = -s_2) \\ = \frac{1}{\sigma\sqrt{d}} \sum_{l \in \mathbb{Z}} \exp\left(-\frac{\pi d}{\sigma^2} \left(l + \frac{v - s_2}{d}\right)^2\right), \quad (22)$$

which means that $p_2(v, s_2)$ is purely a function of $(v - s_2)/d$ with the same concentration behavior as $p_1(u, s_1)$. With the only difference being in the sign of the syndrome these distributions behave exactly the same. Putting both of these together, we have a full characterization of the event of the error $\hat{X}^u \hat{Z}^v$ if we obtain the total syndrome s_1, s_2 which we denote by the probability distribution

$$p(u, v, s_1, s_2) = p_1(u, s_1)p_2(v, s_2). \quad (23)$$

Once the values of s_1 and s_2 are measured and we correct by the smallest displacement, we are left with Pauli noise on the qudit which has an error distribution of $p(u, v|s_1, s_2)$ for $\hat{X}^u \hat{Z}^v$. This can be equivalently modeled by a qudit noise channel which also outputs $|s_1, s_2\rangle \in \mathcal{S}$ in an output register described by a Hilbert space $\mathcal{S}$ where $\langle s'_1, s'_2 | s_1, s_2 \rangle = \delta(s_1 - s'_1)\delta(s_2 - s'_2)$ which just ensures that we can exactly measure the values of s_1 and s_2 . This is the same as the subsystem decomposition used for describing GKP states as a tensor product between logical Hilbert space and the stabilizer eigenvalues which can be used to define a complete basis for the bosonic mode [28,29,38]. We can use this to define an effective logical

noise channel $\mathcal{N}_{\text{logical}} : \mathcal{L}(\mathcal{H}_d) \rightarrow \mathcal{L}(\mathcal{H}_d \otimes \mathcal{S})$ defined as

$$\mathcal{N}_{\text{logical}}(\cdot) = \sum_{u,v} \int ds_1 ds_2 p(u, v, s_1, s_2) \\ \hat{X}^u \hat{Z}^v (\cdot) (\hat{X}^u \hat{Z}^v)^\dagger \otimes |s_1, s_2\rangle \langle s_1, s_2| \quad (24)$$

which once we measure out the extra register reduces to qudit Pauli noise with distribution $p(u, v|s_1, s_2)$. Since the extra register is effectively classical in the sense that it can only give classical information which is the syndrome itself, this noise channel is a classical mixture of Pauli noise with an auxiliary system. We will now examine an achievable rate for this channel which fundamentally makes use of the analog information.

B. Achievable rate using analog information

For a qudit Pauli noise channel with noise distribution $p_{u,v}$ for the error $\hat{X}^u \hat{Z}^v$, we can find the coherent information for the input of a maximally mixed state to be

$$I_c(\mathbb{I}/d, \mathcal{N}_{\text{Pauli}}) = \log_2(d) - H_2(p_{u,v}), \quad (25)$$

where $H_d(p_{u,v}) = -\sum_{u,v} p_{u,v} \log_d(p_{u,v})$. This gives the Hashing bound for Pauli channels which is achievable through the use of random CSS codes [43] hence is necessarily below the capacity of the Pauli channel [44]. Similarly for a classical mixture of Pauli (CMP) channels with an auxiliary system defined as

$$\mathcal{N}_{\text{CMP}}(\cdot) = \sum_{u,v,j} p_{u,v,j} \hat{X}^u \hat{Z}^v (\cdot) (\hat{X}^u \hat{Z}^v)^\dagger \otimes |j\rangle \langle j|, \quad (26)$$

taking the input to be the maximally mixed state yields the coherent information to be

$$I_c(\mathbb{I}/d, \mathcal{N}_{\text{CMP}}) = \log_2(d) - \sum_j p_j \log_2(p_j) \\ + \sum_{u,v,j} p_{u,v,j} \log_2(p_{u,v,j}) \\ = \log_2(d) - H_2(u, v|j) \quad (27)$$

where $H_d(u, v|j) = H_d(p_{u,v,j}) - H_d(p_j)$ is the conditional information entropy of u, v conditioned on j and $p_j = \sum_{u,v} p_{u,v,j}$. If the register containing information on j was traced out, we obtain a Pauli channel with distribution $p_{u,v}$ for which the coherent information will be less than $I_c(\mathbb{I}/d, \mathcal{N}_{\text{CMP}})$ by the mutual information $I(u, v; j)$ which is always non-negative.

From this we can see that previous methods for calculating achievable rates associated to the square GKP qudit [11,22] were in essence calculating the rate associated to the CMP obtained by not having access to the auxiliary system for the channel in Eq. (24). Hence we now define an achievable rate associated to square GKP qudits as $I_c(\mathbb{I}/d, \mathcal{N}_{\text{logical}})$ which equals

$$I_{d,\text{analog}}^{\text{sq}} = \log_2(d) + 2 \int ds_1 \sum_{u=0}^{d-1} p_1(u, s_1) \log_2 \left(\frac{p_1(u, s_1)}{p(s_1)} \right), \quad (28)$$

where we use the fact that $p(u, v, s_1, s_2)$ fully separates into $p_1(u, s_1)p_2(v, s_2)$ and $H(u|s_1) = H(v|s_2)$. These are evaluated for different values of underlying noise strength σ and plotted in Fig. 2(a) in comparison to the coherent information of the Pauli noise channel with probability distribution $p(u, v)$ obtained by throwing out syndrome information. In the numerical evaluation of $I_{d,\text{analog}}^{\text{sq}}$ we observe that the threshold value of σ beyond which the channel does not admit a rate seems independent of the value of d . This threshold value can be numerically checked to be approximately $1/\sqrt{e}$ which has been observed in concatenated GKP schemes that use analog information [23,45].

As d is increased, there seems to be a clear convergence of the value that $I_{d,\text{analog}}^{\text{sq}}$ takes independent of d . In particular we find that for any given σ , once $d \gg \sigma^{-2}$, the

value

$$\left| I_{d,\text{analog}}^{\text{sq}} - \log_2 \left(\frac{1}{\sigma^2 e} \right) \right| \leq \mathcal{O}(e^{-d\pi\sigma^2}) + \mathcal{O}(d^{1/2}\sigma^{-1}e^{-\frac{\pi d}{9\sigma^2}}), \quad (29)$$

the derivation of which is detailed in Appendix D 1. This is an incredibly interesting consequence since it is known from Refs. [10,11] that the coherent information of the Gaussian displacement noise channel $\mathcal{N}_{\text{displ}}$ [see Eq. (14)] is

$$I_c(\mathcal{N}_{\text{displ}}) = \log_2 \left(\frac{1}{\sigma^2 e} \right). \quad (30)$$

Hence if we choose a large enough d and concatenate it to some outer code capable of achieving $I_{d,\text{analog}}^{\text{sq}}$, we can get very close to $I_c(\mathcal{N}_{\text{displ}})$ as an achievable rate. In fact by taking $d \rightarrow \infty$ we can concretely claim $I_c(\mathcal{N}_{\text{displ}})$ as achievable using this particular choice of codes.

The reason for this convergence is roughly the same as the reason $p_1(u, s_1)$ converges to $p_{\text{lim}}(u, s_1)$. We offer a more heuristic description of the convergence of $p_1(u, s_1)$ to $p_{\text{lim}}(u, s_1)$ in Fig. 2(b). The distribution of $p_{\text{lim}}(u, s_1)$ implies that on increasing d , only the contributions of X^u errors with $|u| < \mathcal{O}(\sqrt{d})$ become relevant since the other probabilities are exponentially suppressed. This means that even though the Hilbert space dimension has increased by increasing d , the size of the typical error set is only growing as a function of $\sqrt{d}$. The displacements along $\hat{q}$ that

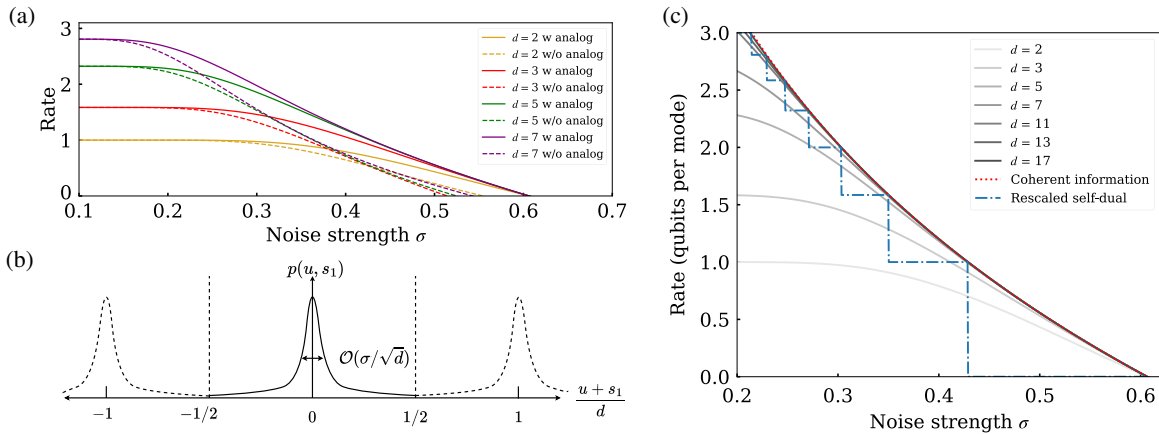


FIG. 2. (a) Achievable rates for the square GKP qudit with analog information ($I_{d,\text{analog}}^{\text{sq}}$) and without analog information ($I_{d,\text{no analog}}^{\text{sq}}$) plotted in solid and dashed lines respectively for different values of d . The dashed line uses the distribution $p(u)$ obtained by averaging over the syndrome hence resulting in not using any analog information. (b) A rough depiction of what the concentration of $p_1(u, s_1)$ looks like in the limit of large d [see Fig. 1(c) for an exact numerical evaluation of $p_1(u, s_1)$ for different values of d]. While there is a contribution from terms that are centered around $\pm l$ for $l = 1, 2, \dots$ (see plotted Gaussians with dashed lines), their contribution is exponentially small in magnitude. Additionally, the spread of $p_{\text{lim}}(u, s_1)$ being $\mathcal{O}(\sigma/\sqrt{d})$ over the normalized variable $(u + s_1)/d$ results in the size of typical error set growing much smaller than the actual dimension of Hilbert space. (c) Numerical evaluation of $I_{d,\text{analog}}^{\text{sq}}$ (for values of $d = 2$ to $d = 17$ shown in solid lines) showing a clear convergence to the coherent information of the Gaussian displacement noise channel (dotted line). Notably for finite and reasonably small d (< 10), we already see the value of $I_{d,\text{analog}}^{\text{sq}}$ get numerically close to coherent information while also crossing the value of rates achieved by rescaled self-dual lattices (dashed-dotted line).

result in the same syndrome are spaced apart by $\sqrt{2\pi}/d$ and as a result increasing d ends up making all the syndromes equally likely. As a result, the marginal distribution $p(s_1)$ approaches the uniform distribution over s_1 between $-1/2$ to $1/2$ differing only in $\mathcal{O}(e^{-\pi d \sigma^2})$. However, it is clear that even in this limit, the syndrome is providing useful information since $p_1(u, s_1)$ is a function of $(u + s_1)/d$. The convergence toward p_{lim} lets us analytically approximate the value of $I_{d,\text{analog}}^{\text{sq}}$ while bounding the difference from the actual values in functions that are exponentially suppressed in increasing the value of d/σ^2 . Putting all of this together, we reach our first claim that through the concatenation of appropriate outer code with square GKP qudits, the coherent information is indeed an achievable rate for all values of the noise strength σ .

Numerically, we note that since the difference in rates is exponentially suppressed in growing d , even for $d = 5$ and $\sigma > 0.4$, the achievable rate is numerically very close to $I_c(\mathcal{N}_{\text{displ}})$. Numerical evaluations of the values of $I_{d,\text{analog}}$ are plotted in Fig. 2(c). In the following section we will show how quantum polar codes can be shown to achieve $I_{d,\text{analog}}^{\text{sq}}$ which shows that we can explicitly construct the GKP codes that have rates very close to $I_c(\mathcal{N}_{\text{displ}})$.

IV. POLAR CODES AS A CANDIDATE FOR CONCATENATION

A. Channel polarization for classical noise

We introduce a notation of u_1^N which represents a vector of length N with elements labeled as u_i , and a subset of it u_i^j is the terms of indices between i to j (inclusive of both). We now motivate the definition of a discrete memoryless channel (DMC). A classical noise channel W can be seen as a map between symbols in an input alphabet $\mathcal{X}$ to an output alphabet $\mathcal{Y}$ where the probability symbol x maps to y is given by $W(y|x)$. The memoryless aspect of this noise channel means that if this were to act individually over N different symbols x_1^N where each $x_i \in \mathcal{X}$, the output symbol y_i only depends on x_i and has no memory of the symbols that came before it. Hence the noise acts independently on each symbol. Common discrete memoryless channels include erasure channels, bit-flip channels and also the additive white Gaussian noise (AWGN) channel. Here we are concerned with dits so $\mathcal{X} \equiv \mathbb{F}_d$ and we will not place any limitations on what the output alphabet $\mathcal{Y}$ is. One can always find the symmetric capacity of this channel which is given by

$$I(W) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{d} W(y|x) \log_d \left(\frac{W(y|x)}{\sum_{x' \in \mathcal{X}} \frac{1}{d} W(y|x')} \right),$$

and will always lie between 0 and 1. This represents the rate (in dit per channel use) of information transfer that

can be achieved through this channel assuming all inputs to be equally likely.

Let us suppose we have a linear transform $G_N : \mathbb{F}_d^N \rightarrow \mathbb{F}_d^N$ and it acts on a dit-string of length N which we represent as u_1^N , and maps it to x_1^N . We now communicate x_1^N over N copies of the noise channel W which maps it to the output y_1^N . Consider the task of figuring out u_1^N using the output y_1^N aided with a genie which supplies the values u_1^{i-1} when we wish to estimate u_i . This by itself can be understood by treating the whole set y_1^N, u_1^{i-1} as an output for the dit u_i whilst all of u_{i+1}^N can take any possible value. Note that this is nothing but a discrete memoryless channel of its own mapping u_i to the tuple (y_1^N, u_1^{i-1}) with some probability $W_N^{(i)}((y_1^N, u_1^{i-1})|u_i)$. Wherever we use the notation $W_N^{(i)}$, we will be referring to the channel $W_N^{(i)}((y_1^N, u_1^{i-1})|u_i)$. The channel $W_N^{(i)}$ has a very interesting property of its own. As the block-length N increases, the fraction of channels that have $I(W_N^{(i)}) \in (1 - \delta, 1]$ begins to tend to $I(W)$ whilst the fraction of channels with $I(W_N^{(i)}) \in [0, \delta)$ tends to $1 - I(W)$ for any $\delta > 0$ for a particular choice of G_N , namely the polar transform [described in Eq. (35)] [46,47]. In essence, all these channels either are nearly perfect for information transfer or are completely useless which is where the term polarization arises from.

Another relevant channel parameter is the Bhattacharya distance $Z(W)$ defined as

$$Z(W) = \frac{1}{d(d-1)} \sum_{x, x' \in \mathcal{X}, x \neq x'} \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x')}, \quad (31)$$

which also happens to give an inequality for the capacity as

$$\begin{aligned} \log_d \left(\frac{d}{1 + (d-1)Z(W)} \right) &\leq I(W) \\ &\leq 1 - \frac{(1 - \sqrt{1 - Z(W)^2})}{\log_2(d)}, \end{aligned} \quad (32)$$

proven in Ref. [46]. The importance of this is that $I(W) \rightarrow 1$ if and only if $Z(W) \rightarrow 0$. This has a direct consequence in the way we design a polar code since the values of $Z(W)$ can be estimated through Monte-Carlo sampling [48]. Note that $\sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x')} = \sum_{y \in \mathcal{Y}} \sqrt{\frac{W(y|x)}{W(y|x')}} W(y|x')$ and $W(y|x')$ is the output probability distribution for y assuming the input to be x' , and so we observe that

$$Z(W) = \frac{1}{d(d-1)} \sum_{x, x' \in \mathcal{X}, x \neq x'} \left\langle \sqrt{\frac{W(y|x)}{W(y|x')}} \right\rangle_{y|x'}, \quad (33)$$

where the average $\langle \cdot \rangle_{y|x'}$ is for the random variable y sampled over distribution $W(y|x')$. Given that the decoding

procedure for polar codes involves calculating the values of $\frac{W(y|x)}{W(y|x')}$ (see Appendix A 3), we can estimate the value of $Z(W)$ through Monte-Carlo sampling. This does have its drawbacks since these values are very close to zero and come from a heavily tailed distributions making Monte-Carlo methods somewhat unfaithful for exact estimation although useful for identification of whether $Z(W) \ll 1$ or $Z(W) \approx 1$. These values can be more accurately tracked using density evolution methods [49].

A polar code is defined in terms of the set of symbols that are chosen to be frozen before the encoding procedure. The set $\mathcal{A} \subseteq \{1, \dots, N\}$ is the set of symbols that are used for the encoding procedure. The error probability of a polar code under a successive cancellation (SC) decoder can be upper bounded by

$$P_e \leq (d-1) \sum_{i \in \mathcal{A}} Z(W_N^{(i)}), \quad (34)$$

which follows from the fact that the SC decoder is sequentially doing the maximum likelihood estimate of u_i using $W_N^{(i)}$, a task for which the error probability is bounded above by $(d-1)Z(W_N^{(i)})$ [46]. This is notably not the same as the overall maximum likelihood decoding since the channels $W_N^{(i)}$ only use knowledge of the symbols u_1^{i-1} and assume even the possibly frozen symbols in u_{i+1}^N to be random. The asymptotic behavior of $Z(W_N^{(i)})$ is proven in Ref. [50] where it is shown that the fraction of $W_N^{(i)}$ with $Z(W_N^{(i)}) \leq 2^{-N^\beta}$ for any $\beta < 1/2$ approaches the value of $I(W)$ as $N \rightarrow \infty$. This hence gives a sequence of codes with $P_e \rightarrow 0$ and $|\mathcal{A}|/N \rightarrow I(W)$ hence being capacity achieving. For a fixed N , we refer to the “good” indices i as those with $Z(W_N^{(i)}) < \delta$ and the “bad” indices to be everything else. The good indices will go in $\mathcal{A}$, hence having u_i carry information and all the other u_i with $i \in \mathcal{A}_c$ (the complement of $\mathcal{A}$) will be frozen to a particular symbol. We can choose a sufficiently small δ to ensure that the channels $W_N^{(i)}$ are sufficiently good enough for information transmission. Construction of polar codes is done by identifying a way to order the values of $Z(W_N^{(i)})$ for the channels, which is not an easy task since exact evaluation of these parameters requires tracking an exponentially growing output space. As such, there have been various methods devised for ordering the Bhattacharya distances of each channel $W_N^{(i)}$ [51–53].

We refer the reader to Appendix A 1 for a description of the encoding procedure and Appendix A 3 for the SC decoding procedure for a polar code. We make use of the fact that polarization is known to happen for channels with input alphabets that are prime in number [46] as well as the fact that the kernel used for the encoding operation can be chosen to be any matrix of the form

$$G_N = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}^{\otimes n}, \quad (35)$$

where $N = 2^n$, $G_N \in \mathbb{F}_d^{N \times N}$, and $\alpha \in \mathbb{F}_d \setminus \{0\}$ [54]. We will refer to α as the kernel choice. We now proceed to show how the principle of channel polarization can be used to achieve the rate $I_{d,\text{analog}}^{\text{sq}}$.

B. CSS-like entanglement assisted polar code construction

Given their capacity achieving properties in the case of classical memoryless noise channels, it was only natural that there was work done to extend this to quantum noise channels. Wilde and Guha showed that quantum polar codes can be constructed for achieving the capacity of classical-quantum channels [55] as well as the coherent information of degradable channels with classical environment [55], albeit without an efficient decoding method. Dupuis *et al.* [21] observed that a single qubit Pauli noise channel can be treated as a combination of two DMCs with one representing bit-flip noise and the other representing phase-flip (conditioned on whether a bit-flip has occurred or not), a CSS-like entanglement assisted code can be constructed using two classical polar codes. This code achieves a net coding rate equal to the Hashing bound of this noise channel. Importantly this scheme has both efficient encoding and decoding since the encoding circuit consist of just CNOT operations and the decoder simply reverses the encoder and then the syndrome can be fed into a classical decoder. This was further generalized to beyond Pauli channels in Refs. [56,57] showing a channel splitting and recombining method using random two-qubit Clifford operations. These have also been extended for qudit input channels as shown in Ref. [58]. We will mainly be using the principles as they are presented in Ref. [21] and extending them for qudit CMP channels. We refer the reader to Ref. [59] for a more detailed treatise on quantum polar codes.

We first begin by examining what the polar transform looks like over qudits. We wish to be able to map all the N qudit states in the diagonal basis for $\hat{Z}$ (also chosen as the computational basis) labeled as $|\mathbf{z}\rangle$ ($\mathbf{z} \in \mathbb{F}_d^N$) to $|G_N \mathbf{z}\rangle$. The diagonal basis for $\hat{X}$ will be represented by $|\tilde{\mathbf{x}}\rangle$ ($\mathbf{x} \in \mathbb{F}_d^N$). Since a single unit of the encoding operation takes $(z_1, z_2) \rightarrow (z_1 + \alpha z_2, z_2)$ for $z_1, z_2 \in \mathbb{F}_d$, this operation is equivalent to α repetitions of the CSUM gate. Extending this to the quantum case we can define the unitary

$$\hat{V} = \sum_{\mathbf{z} \in \mathbb{F}_d^N} |G_N \mathbf{z}\rangle \langle \mathbf{z}| = \sum_{\mathbf{x} \in \mathbb{F}_d^N} |G_N^{-T} \tilde{\mathbf{x}}\rangle \langle \tilde{\mathbf{x}}|, \quad (36)$$

where we can note that since V is obtained by the use of $\mathcal{O}(N \log(N))$ CSUM gates, each gate in the X basis

becomes a CSUM⁻¹ gate with the control and target swapped. Hence effectively the polar transform in the $\hat{X}$ basis is given by

$$G_N^{-T} = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}^{\otimes n}, \quad (37)$$

which albeit is a different kernel, it will still give channel polarization as shown in Ref. [54]. It should also be noted that the order of the dits is effectively reversed in the x basis. As was observed in Refs. [21,58], the qudit Pauli noise channel with $p_{u,v}$ probability for the operation $\hat{X}^u \hat{Z}^v$ can be understood as two DMCs

$$\begin{aligned} W_A(z + u|z) &= \sum_v p_{u,v}, \\ W_P((x + v, u)|x) &= p_{u,v}, \end{aligned} \quad (38)$$

where W_A represents bit-flip noise and W_P represents phase-flip noise. Notably these are not independent noise channels in general and so to account for that, we represent the phase noise as a DMC which also outputs the dit u indicating the amount of dit-flip that has occurred already since this conditions the probability of a phase-flip after that. We now consider the classical polar codes that can achieve the capacity of $I(W_A)$ and $I(W_P)$ in dit per channel use [or equivalently $\log_2(d)I(W_A)$ and $\log_2(d)I(W_P)$ in bit per channel use] and note that we can simultaneously achieve their capacity since the polar transform $\hat{V}$ in Eq. (36) does the polar transform G_N in the Z basis and the polar transform G_N^{-T} in the X basis. To be able to communicate quantum information through a particular index $i \in \{1, \dots, N\}$, we require it to be both a “good” index against bit-flip and phase-flip noise. Since the polarization in the phase basis occurs in the reversed order of indices, this means we would always have to consider four disjoint sets which an index i must lie in which are as follows.

- (1) $\mathcal{I}$: set of indices good against both bit and phase-flip noise.
- (2) $\mathcal{A}$: set of indices good against phase-flip noise but bad against bit-flip noise.
- (3) $\mathcal{P}$: set of indices good against bit-flip noise but bad against phase-flip noise.
- (4) $\mathcal{E}$: set of indices bad against both noise.

Naturally this means that we can use qudits with index $i \in \mathcal{I}$ to send quantum information. Qudits with index $i \in \mathcal{A}$ will be required to be frozen in the Z basis and qudits with index $i \in \mathcal{P}$ will be required to be frozen in the X basis. Notably due to the structure of $\hat{V}$ being only CSUM gates, $\hat{V}^\dagger \hat{X}_i \hat{V}$ and $\hat{V}^\dagger \hat{Z}_i \hat{V}$ always stay as $\hat{X}$ or $\hat{Z}$ strings. Since $\mathcal{A}$ and $\mathcal{P}$ are non-intersecting, freezing their indices will also make sure we end up with commuting stabilizers which are either purely made of Pauli X or Pauli Z operations giving

a CSS like construction. The complication arises when we consider $\mathcal{E}$ which must be frozen in both X and Z bases simultaneously.

Using an entanglement assisted coding method [60–62], these can then be chosen to be shared Bell pairs between the sender and receiver which would make this a valid coding method. The two qudit state which is simultaneously stabilized by $\hat{X}_A^\dagger \hat{X}_B$ and $\hat{Z}_A^\dagger \hat{Z}_B$ is a Bell state where the qudit with B can be understood as frozen in both $\hat{X}$ and $\hat{Z}$ basis since the joint measurements of these stabilizers will have a fixed outcome. We consider the sender and receiver to be sharing Bell pairs that would consist of these frozen qudits that can now be considered as frozen in both the X and Z bases. As such, these Bell pairs are fully consumed within this coding scheme and hence this is an example of an entanglement assisted scheme. The set $\mathcal{A} \cup \mathcal{E}$ would be frozen in the Z basis (hence $|\mathcal{A} \cup \mathcal{E}| \approx N(1 - I(W_A))$) and the set $\mathcal{P} \cup \mathcal{E}$ would be frozen in the X basis (hence $|\mathcal{P} \cup \mathcal{E}| \approx N(1 - I(W_P))$), and here $\mathcal{E}$ is frozen in both bases by the fact that it is shared Bell pairs. As a result of the possible entanglement assistance, the net coding rate (in qudit per channel use) this achieves is given by

$$\frac{|\mathcal{I}| - |\mathcal{E}|}{N} \rightarrow I(W_A) + I(W_P) - 1 = 1 - H_d(p_{u,v}), \quad (39)$$

where we detail the encoding and decoding of this in Appendix B. Here we account for the fact that this scheme makes use of entanglement assistance by defining the net rate to be $(|\mathcal{I}| - |\mathcal{E}|)/N$ which accounts for the fact that this consumes $|\mathcal{E}|$ Bell pairs for the communication.

We describe the structure of our quantum polar code concatenation with single-mode GKP qudits in Fig. 3. Since we are working with a classical mixture of Pauli noise Eq. (24), we can consider the syndrome to be part of the noisy output. Since we are using a single mode square GKP code undergoing an isotropic displacement noise, the logical $\hat{X}$ and $\hat{Z}$ errors are uncorrelated, which is not generally true for all single mode GKP codes (we explore this in Appendix D3). On including the syndrome as part of the noisy output, we obtain two noise channels

$$\begin{aligned} W_1(z + u, s_1|z) &= p_1(u, s_1), \\ W_2(x + v, s_2|x) &= p_2(v, s_2), \end{aligned} \quad (40)$$

that are still DMC, and will obey all the polarization requirements that W_A and W_P do. Hence we similarly see that

$$\frac{|\mathcal{I}| - |\mathcal{E}|}{N} \rightarrow I(W_1) + I(W_2) - 1 = \frac{I_{d,\text{analog}}^{\text{sq}}}{\log_2(d)}, \quad (41)$$

which essentially means that if we concatenate a quantum polar code constructed using the knowledge of noise channels W_1 and W_2 , we can obtain an explicit concatenated

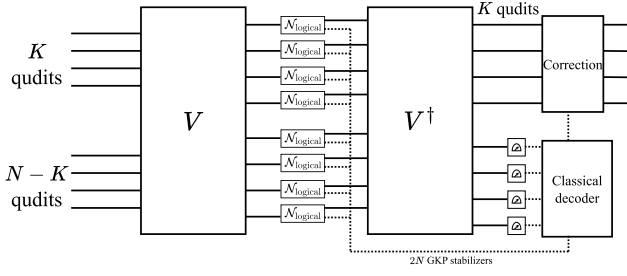


FIG. 3. A heuristic circuit diagram explaining the structure of the concatenation. Here we treat each single mode square GKP as its own qudit which undergoes an encoding operation V which creates the codewords for the concatenated outer polar code. The operations in V are logical qudit operations (in this case purely composed of CSUM gates in the case of the polar code). The K qudits are used for encoding information and the $N - K$ qudits are frozen to some chosen state. The logical noise operation $\mathcal{N}_{\text{logical}}$ acts on all the encoded qudits and also outputs classical information in the form of the analog GKP syndrome of each individual mode. Following this the decoding operation proceeds, which involves inverting the encoding operation and then measuring the previously frozen $N - K$ qudits which gives the syndrome information of the outer polar code. This along with all the individual GKP syndromes is then used in a classical decoder (the successive cancellation decoder) giving a correction to be applied on the K qudits carrying information.

square GKP code capable of achieving $I_{d,\text{analog}}^{\text{sq}}$. The success of the code relies on a decoding task which is the combination of decoding the bit and phase errors separately. Hence we must consider the failure probabilities of both decoders

$$\begin{aligned} P_{e,1} &\leq (d-1) \sum_{i \in \mathcal{I} \cup \mathcal{P}} Z((W_1)_N^{(i)}), \\ P_{e,2} &\leq (d-1) \sum_{i \in \mathcal{I} \cup \mathcal{A}} Z((W_2)_N^{(i)}), \end{aligned} \quad (42)$$

which must both go to zero, which is ensured by the individual polarization of W_1 and W_2 . Here the channel $(W_1)_N^{(i)}$ is shorthand for $(W_1)_N^{(i)}((y_1^N, u_1^{i-1})|u_i)$ which maps input dit u_i to the output of (y_1^N, u_1^{i-1}) assuming all symbols u_{i+1}^N to be equally likely similar to the standard definition we use for $W_N^{(i)}$ in Sec. IV A. However, the channel $(W_2)_N^{(i)}$ is shorthand for $(W_2)_N^{(i)}((y_1^N, u_{i+1}^N)|u_i)$ which maps input dit u_i to output of (y_1^N, u_{i+1}^N) (where y_1^N is the output of N copies of W_2 receiving the encoded dit-string $x_1^N = G_N^{-T} u_1^N$). This is to account for the fact that for the noise channel W_2 , the dits will be decoded in a reversed order with the same successive cancellation decoder. It is important to point out that we only make this claim for prime values of d since this result hinges on the ability that arbitrary kernels can be used for polarization as well as the polarization sharply splits channels into having $I(W) \rightarrow 1$ or $I(W) \rightarrow 0$, which is not necessarily true for all kernels if non-prime values

of d were to be used [46]. It will be interesting to explore the generalization of these results to arbitrary d , which we leave for future work.

C. Numerical results and discussion

Designing a polar code requires being able to find the values of $Z(W_N^{(i)})$ to be able to judge which indices are good and bad for information communication. In general, this is not an easy task since it involves summing over an exponentially increasing number of terms as can be observed from the fact that the output alphabet of $W^{(i)}$ lies in $\mathcal{Y}^N \times \mathcal{X}^{i-1}$. This hence calls for a sampling-based method to be able to estimate $Z(W^{(i)})$. We note that the channels we are working with are symmetric, which is to say that for any $a \in \mathbb{F}_d$, there is a permutation $\pi_a : \mathcal{Y} \rightarrow \mathcal{Y}$ such that $W(y|x) = W(\pi_a(y)|x \oplus a)$. This is trivially true for the channels W_1 and W_2 in Eq. (40). Due to this symmetry, we only need to be concerned with the effect of these channels on the all 0 input and being able to decode it back to all 0 perfectly. Also, this makes the choice of frozen values to not have any consequence on code performance, hence we can choose the frozen values u_{A_c} to always be 0. Due to this symmetry, we note that $Z(W)$ depends only on the values of $\langle \sqrt{W(y|a)/W(y|0)} \rangle_y$ averaged over all possible outputs of W . The SC decoder in fact finds these values for the channels $W^{(i)}$ during the decoding procedure which it then uses to make the decision of what the best estimate $\hat{u}_i$ is for the symbol u_i . The closer this expression is to 0, this reflects more sureness for $u_i = 0$ which is the correct value, and the closer this is to 1 reflects less sureness. Hence we pick M samples from the distribution the outputs follow and then in $\mathcal{O}(MN \log(N))$ time we are able to estimate all the values of $Z(W^{(i)})$ with an absolute error of $\mathcal{O}(M^{-1/2})$ using Monte-Carlo estimation. Due to the symmetries in the definition of W_1 and W_2 , we will have $Z((W_1)_N^{(i)}) = Z((W_2)_N^{(N-i)})$ hold for all $i \in \{1, \dots, N\}$ even though they effectively have different kernels of G_N and G_N^{-T} respectively. This is due to the fact that the channel W_1 itself polarizes the exact same way under kernel of α and $-\alpha$ due to a symmetry in the way it acts on input dit of $+x$ and $-x$ which we detail in Appendix E 1. We require both $Z(W_1^{(i)})$ and $Z(W_2^{(i)})$ to simultaneously approach 0 for i to be a good index. We can see channel polarization in action in Fig. 4 which shows the upper bounds on $Z(W_1^{(i)})$ and $Z(W_2^{(i)})$ for $N = 512$ to 4096 for a square qudit ($d = 5$) with $\sigma = 0.4$ and kernel choice $\alpha = 2$ and $\alpha = 1$ in blue and orange respectively. The fraction of indices satisfying both $Z(W_1^{(i)}) < \delta$ and $Z(W_2^{(i)}) < \delta$ can be clearly seen to be increasing with N as one would expect for both choices of α .

In this work we will be relying on this heuristic method which uses Monte-Carlo methods for figuring out the relative ordering of $Z((W_1)_N^{(i)})$ in an approximate manner for

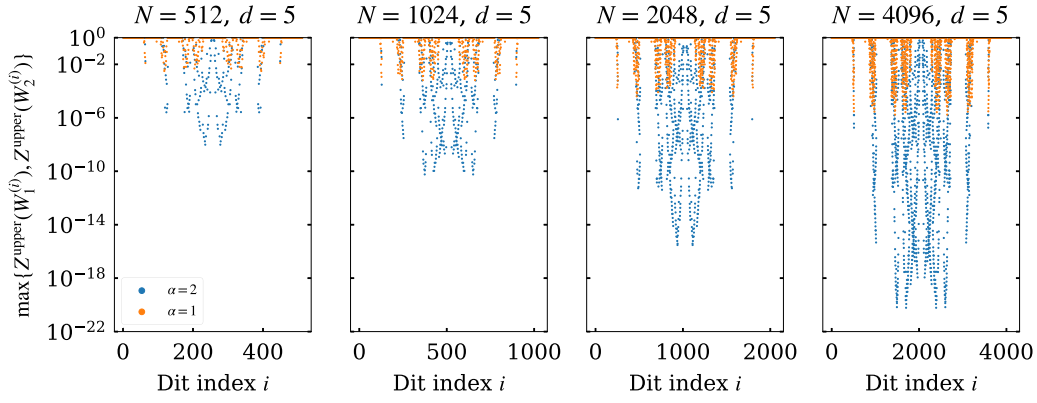


FIG. 4. Effects of channel polarization on increasing block length N from 2^9 to 2^{12} . We obtain upper bounds on the values of $Z((W_1)_N^{(i)})$ and $Z((W_2)_N^{(i)})$ for displacement spread $\sigma = 0.4$ using kernel with $\alpha = 2$ (blue dots) and $\alpha = 1$ (orange dots) for $d = 5$. These upper bounds are labeled as $Z((W_1)_N^{(i),\text{upper}})$ and $Z((W_2)_N^{(i),\text{upper}})$. Information will be encoded in dits that have both $Z((W_1)_N^{(i)})$ and $Z((W_2)_N^{(i)})$ being small hence we plot $\max\{Z((W_1)_N^{(i),\text{upper}}), Z((W_2)_N^{(i),\text{upper}})\}$. The indices that have both the values being small can be seen to be increasing from $N = 512$ to $N = 4096$. Notably, the polarization phenomenon is observed to be quicker using the kernel of $\alpha = 2$ compared to $\alpha = 1$ as can be inferred by comparing the fraction of indices satisfying $\max\{Z((W_1)_N^{(i),\text{upper}}), Z((W_2)_N^{(i),\text{upper}})\}$ for some small enough δ .

our constructions for the sake of proof of concept. We attempt to construct sequences of polar codes satisfying

$$P_{e,1} \leq c_e N^{-\beta}, \quad P_{e,2} \leq c_e N^{-\beta}, \quad (43)$$

for a particular value of noise strength σ and constants $0 < c_e < 1$ and $\beta > 0$. We do this by finding the best choices of $\mathcal{I}, \mathcal{A}, \mathcal{P}$ and $\mathcal{E}$ based on the estimated values of $Z((W_1)_N^{(i)})$ and $Z((W_2)_N^{(i)})$ motivated by Eq. (42). This means identifying the indices $Z((W_1)_N^{(i)}) \leq c_e N^{-(1+\beta)}$ which can be seen as a hypothesis test which requires $M = \Omega(N^{2(1+\beta)})$ for guaranteeing high confidence. However this is not very tractable since we require large blocklengths N that can demonstrate channel polarization in effect. Based on the martingale properties of $Z((W_1)_N^{(i)})$, the fraction of indices i that have $Z((W_1)_N^{(i)}) \in (\delta', 1 - \delta')$ approaches 0 for $\delta' = 2^{-N^{\beta'}}$ for $0 < \beta' < 1/2$. Hence, as long as N is large enough, if we observe that $Z((W_1)_N^{(i)}) < \delta$ for $\delta = \mathcal{O}(1)$ and $\delta < 1 - \delta'$, it is very likely that $Z((W_1)_N^{(i)}) < \delta'$ as well since otherwise it would imply $Z((W_1)_N^{(i)}) \in (\delta', 1 - \delta')$. There will always exist N_0 such that for $N > N_0$, we have $c_e N^{-(1+\beta)} > 2^{-N^{\beta'}}$ for a fixed $c_e, \beta, \beta' > 0$. Hence we would likely have $Z((W_1)_N^{(i)}) < c_e N^{-(1+\beta)}$ on observing $Z((W_1)_N^{(i)}) < \delta$ for a large enough N . This heuristically explains why even for a reasonable $M = 2048$ sample we are able to construct polar codes that perform reasonably well although they are not guaranteed to be the best possible constructions.

To actually guarantee that the sequence of codes we obtain has an error probability that is upper bounded by $c_e N^{-\beta}$, we estimate the error probability for the explicitly constructed codes we obtain using Monte-Carlo sampling.

More details on the numerical results for the error probabilities can be found in Appendix E. We report our results for different values of d in Fig. 5. For a fixed value of σ each sequence is plotted in the same legend as N is varied from 2^9 to 2^{18} . The rate $(|\mathcal{I}| - |\mathcal{E}|)/N$ is strictly increasing in N (as one would expect [46]) and can be seen to approach the values of $I_{d,\text{analog}}^{\text{sq}}$. For each of these sequences (in increasing N) of codes at a given value of σ , the total error probability is upper bounded by a function that approaches zero on increasing N while we also observe the rate K/N increasing. Using the asymptotic behavior of the Bhattacharya distances, we are guaranteed to be able to find a code that has an error probability below this decreasing function while the rate K/N approaches $I(W)$ as long as we have exact knowledge of each $Z(W_N^{(i)})$.

We also note that the choice of kernel has a clear impact on performance. We find that for $d = 5$, the kernel choice of $\alpha = 2$ is far better than that of $\alpha = 1$ in the sense that the polarization occurs more dramatically at the same value of N as can be noted in Fig. 4. It is worth noting that there are not direct distance scaling guarantees for polar codes and in large part they are not a direct measure of performance. Using the upper bound in Eq. (42), the error probabilities converge to zero for large enough N (provably upper bounded by $N 2^{-N^{\beta}}$ for $0 < \beta < 1/2$ [50]) which is essential in showing their capacity achieving properties. This also is sufficient to claim good performance without delving into distance since the error probability is ultimately the relevant metric for performance. In their application in classical settings, their distance can be improved by a careful selection of the information set [63] if necessary, since the distance equals $2^{\min_{i \in \mathcal{A}} (wt(i-1))}$ where $wt(i-1)$ is the Hamming weight of the binary representation of the

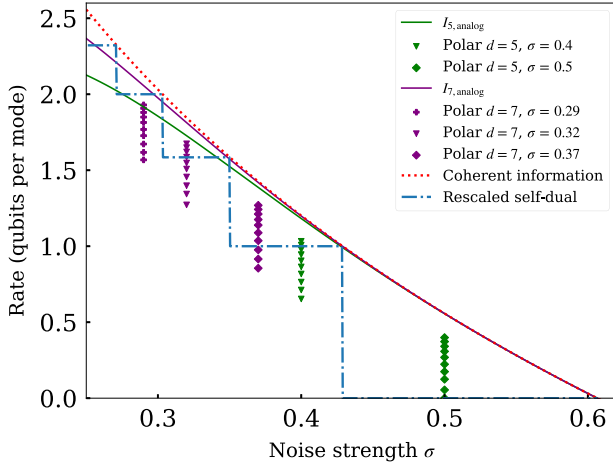


FIG. 5. Performance of polar codes concatenated to square GKP qudits of $d = 2$, $d = 5$, and $d = 7$. Here we find an explicit sequence of codes with increasing N which satisfies $P_{e,1} \leq c_e N^{-\beta}$ and $P_{e,2} \leq c_e N^{-\beta}$ at a given value of σ as the value of N is increased (see Appendix E for a numerical verification of this). For the above plot we use $c_e = 0.5$ and $\beta = 1/4$. The kernel choice for $d = 5$ is $\alpha = 2$, and for $d = 7$ is $\alpha = 3$. We find that we are clearly able to approach the solid line representing $I_{d,\text{analog}}^{\text{sq}}$ for $d = 5$ and $d = 7$ which in turn is able to exceed the rate achievable using rescaled self-dual lattices [11] (dashed-dotted line).

number $i - 1$ for a classical dit polar code of blocklength N encoding K dits with $\mathcal{A}$ being the set of information indices using polar transform with kernel choice α [also denoted by $(N, K, \mathcal{A})_{d,\alpha}$]. This can be seen to be true by simply considering that the encoding procedure of a polar code uses a binary tree structure and remains true for any kernel choice. For the kernel choice of $\alpha = 1$ the smallest weight codeword can also just consist of $2^{\min_{i \in \mathcal{A}} (wt(i-1))}$ 1s and the rest being zeroes. For a bit-flip noise like W_1 where usually $p(1, s_1)$ is much larger than any other $p_1(u, s_1)$ [except for $p(0, s_1)$] this shows how a different kernel choice can lead to the nearest non-zero codeword having smaller probability of being reached from the zero codeword which can then affect finite-blocklength performances.

V. ACHIEVING THE CAPACITY OF PURE-LOSS AND AMPLIFICATION CHANNELS

A. Quantum capacity of the pure-loss and amplification channels

The bosonic pure-loss channel can be modeled as a beam-splitter interaction with the environment which is in the vacuum state following which the environment is traced out. Similarly the amplification channel can be modeled as a two-mode squeezing interaction with the environment in vacuum state following which the environment is traced out. We will consider the loss channel to

have a transmittance $\eta < 1$ and the gain of the amplification channel to be $G > 1$. We can also write these noise channels as follows [17]:

$$\mathcal{N}_{\text{loss}}(\cdot) = \sum_{l=0}^{\infty} \hat{E}_l(\cdot) \hat{E}_l^\dagger, \quad \mathcal{N}_{\text{amp}} = \sum_{l=0}^{\infty} \hat{A}_l(\cdot) \hat{A}_l^\dagger, \quad (44)$$

where the Kraus operators are defined as

$$\begin{aligned} \hat{E}_l &= \left(\frac{1-\eta}{\eta} \right)^{l/2} \frac{\hat{a}^l}{\sqrt{l!}} \eta^{\hat{n}/2}, \\ \hat{A}_l &= \frac{(G-1)^{l/2}}{\sqrt{l!G}} G^{-\hat{n}/2} (\hat{a}^\dagger)^l, \end{aligned} \quad (45)$$

and so it can be seen that they have a very similar form. Both of these channels are degradable [64], and so their capacity matches the maximal one-shot coherent information [10]. Hence their capacity is given by

$$C_Q(\mathcal{N}_{\text{loss/amp}}) = \max \left(\log_2 \left(\left| \frac{\tau}{1-\tau} \right| \right), 0 \right), \quad (46)$$

where $\tau = \eta$ for loss and $\tau = G$ for amplification. In Ref. [17], due to the similar form of these two channels, it is shown that the infidelities of an infinite energy GKP code after experiencing this noise and then using the near-optimal transpose recovery are upper bounded by $\frac{1}{4} \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus 0} e^{-\pi g |\mathbf{x}|^2}$ [with $g = \eta/(1-\eta)$ for loss and $g = G/(G-1)$ for amplification] where $\mathcal{L}^\perp$ is the dual lattice of the multi-mode GKP code. This upper bound holds true for arbitrary multi-mode GKP codes, which also includes any concatenated square GKP code. Hence, if we find a capacity achieving sequence of codes for the pure-loss channel that make use of the transpose recovery, this same sequence of codes would also achieve the capacity of the amplification channel when using the transpose recovery. Hence for the rest of this discussion, we will show a capacity achieving sequence for the pure-loss since the extension to the amplification will trivially follow.

B. Existence of a capacity achieving sequence of codes

While we have largely discussed square GKP qudits in this work, we now discuss general GKP lattices and proceed to show a capacity achieving sequence of lattices that are obtained through the concatenation of a square GKP qudit with a qudit stabilizer code obtained through a self-orthogonal code in $GF(d^2)$. We use the notation $[N, K]_d$ for a classical linear code which encodes K dits into N dits. We refer the reader to Appendix F for a quick introduction to quantum stabilizer codes obtained using self-orthogonal codes in $GF(d^2)$ where we mainly describe the ideas introduced in the works of Refs. [25, 65, 66]. The set of all possible self-orthogonal $[N, (N-K)/2]_{d^2}$

codes will be denoted by $\mathcal{T}$. This set is useful since we can map each $[N, (N - K)/2]_{d^2}$ code to a valid stabilizer group which is a subset of the N qudit (d level) Pauli group generated by $N - K$ independent stabilizers. The self-orthogonality is the key feature which ensures that this is a valid stabilizer group. The set $\mathcal{T}$ has a notion of balanced-ness [67] which is to say that the number of codes in $\mathcal{T}$ that contain a particular self-orthogonal $\mathbf{a} \in \mathbb{F}_{d^2}^N \setminus \{0\}$ is independent of $\mathbf{a}$ itself. Not all $\mathbf{a} \in \mathbb{F}_{d^2}^N \setminus \{0\}$ are self-orthogonal with $(1/d)(d^{2N} + (d - 1)(-d)^N) - 1$ elements of $\mathbb{F}_{d^2}^N \setminus \{0\}$ satisfying this property [68]. This property is incredibly useful in proving the existence of good codes by making claims on the average of some function taken over the set of codes $\mathcal{T}$. This is made use of in proving the Minkowski-Hlawka theorem for classical codes over prime fields [67], existence of good CSS codes [43] and also existence of good stabilizer codes constructed from self-orthogonal codes in $GF(d^2)$ [25].

An N mode GKP code is obtained through a stabilizer group generated using $2N$ independent displacements. Since all these displacements must necessarily commute, this enforces that the symplectic inner product of two displacements must be an integer multiple of 2π . If we consider displacements defined as $\hat{D}(\xi) = \exp(-\xi^T J \hat{x})$ where $\hat{x} = \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix}$ and J is the symplectic form, we can define the stabilizer lattice $\mathcal{L}$ to be a symplectically integral lattice and the stabilizer group consists of $\hat{D}(\xi)$ for any $\xi \in \mathcal{L}$. The connection of symplectically integral lattices to GKP codes was explored in Refs. [11, 14] and has been extensively studied in Refs. [15, 16] which also explore how gauge choices show up in the lattice formalism. For the lattice $\mathcal{L}$, the GKP code obtained encodes $\det(\mathcal{L})$ levels and the symplectic dual $\mathcal{L}^\perp$ contains all ξ that have $D(\xi)$ be a valid logical GKP operator. A valid GKP lattice must satisfy $\mathcal{L} \subseteq \mathcal{L}^\perp$. Additionally, we can rescale lattices to increase the number of levels it encodes since the lattice $\sqrt{\lambda}\mathcal{L}$ encodes $\lambda^N \det(\mathcal{L})$ levels. This is particularly useful when using self-dual lattices $\mathcal{L}^\perp = \mathcal{L}$ since they are unimodular $\det(\mathcal{L}) = 1$ and hence have to be rescaled by $\sqrt{\lambda}$ to encode information. However for the lattice to remain symplectically integral λ must be an integer. The lattice of the GKP code corresponding to N square qudits is $\sqrt{d}\mathbb{Z}^{2N}$ and so the dual of this is $\mathbb{Z}^{2N}/\sqrt{d}$. Concatenating some $[[N, K]]_d$ code to N square GKP qudits restricts the set of what remains as a valid logical displacement, and so the new effective lattice of the overall N mode code $\mathcal{L}^\perp$ is a sublattice of $\mathbb{Z}^{2N}/\sqrt{d}$ and $\det(\mathcal{L}) = d^K$ which is the logical dimension of this lattice.

We now discuss the infidelity of a GKP code against bosonic pure-loss followed by a transpose recovery channel which is near optimal based on the results from Ref. [17]. From Lemma 15 in Appendix F of Ref. [17], it follows that for any GKP code which is generated using

a symplectically integral lattice $\mathcal{L}$, the infidelity after transpose recovery against pure-loss satisfies

$$\epsilon \leq \frac{1}{4} \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} e^{-\pi \frac{\eta}{1-\eta} |\mathbf{x}|^2}, \quad (47)$$

which provides a very useful upper bound on the performance of a particular GKP code to the pure-loss channel. If we wish to show a rate to be achievable, it is sufficient to find a sequence of lattices $\mathcal{L}_n \subseteq \mathbb{R}^{2n}$ such that

$$\lim_{n \rightarrow \infty} \left(\sum_{\mathbf{x} \in \mathcal{L}_n^\perp \setminus \{0\}} e^{-\pi \frac{\eta}{1-\eta} |\mathbf{x}|^2} \right) \rightarrow 0 \quad \text{and} \quad \frac{\log_2(\det(\mathcal{L}_n))}{n} \rightarrow R. \quad (48)$$

In Ref. [17], it is shown if 2^R is an integer less than equal to $\eta/(1 - \eta)$, it can be achieved through the rescaling of self-dual lattices ($\mathcal{L} = \mathcal{L}^\perp$). This exact limitation also showed up in the achievable rates for the Gaussian displacement noise [11], however as we have shown in this work, that can be surpassed through using concatenated square GKP qudits of large enough dimension. Considering the family of unimodular lattices $\mathcal{L} = \mathcal{L}^\perp$, the Buser-Sarnak theorem [18] showed that for any rotationally invariant $f : \mathbb{R}^{2N} \rightarrow \mathbb{R}$, the following holds:

$$\left\langle \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} f(\mathbf{x}) \right\rangle = \int d^{2N} \mathbf{x} f(\mathbf{x}) \quad (49)$$

where the averaging over the set of all unimodular lattices satisfying $\mathcal{L} = \mathcal{L}^\perp$. This allows for the existence of a particular lattice which has the expression $\sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} f(\mathbf{x}) \leq \det(\mathcal{L}) \int d^{2N} \mathbf{x} f(\mathbf{x})$. This directly shows how much $\sqrt{\lambda}\mathcal{L}$ can be scaled up by while having the sum still upper bounded by something that approaches 0 for $N \rightarrow \infty$. We will refer to such a lattice as a good lattice.

To demonstrate the capacity achieving properties for concatenated square GKP codes, we introduce the following loose definition.

Definition 1 (δ -good lattice). A symplectically integral lattice $\mathcal{L}$ which is weakly self-dual $\mathcal{L} \subseteq \mathcal{L}^\perp$ is δ -good if it satisfies

$$\sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} f(\mathbf{x}) \leq \det(\mathcal{L}) \int d^{2N} \mathbf{x} f(\mathbf{x}) + \delta, \quad (50)$$

for the $f(\mathbf{x}) = e^{-g\pi |\mathbf{x}|^2}$ for any $g > 0$.

Hence now we only need to find a sequence of lattices that are δ -good with $\delta \rightarrow 0$ along this sequence. We find

that under this notion of good-ness we can also define a similarity between a 0-good lattice which is good in the same way that the capacity achieving self-dual lattices are to some δ -good lattice obtained through concatenation.

Consider the set $\mathcal{T}_{\mathcal{L}}$ containing the GKP lattices representing $[[N, K]]_d$ codes obtained by self-orthogonal $[N, (N - K)/2]_{d^2}$ codes concatenated to N modes with each mode having square GKP qudit of d levels. We show that over this set of lattices

$$\left\langle \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} f(\mathbf{x}) \right\rangle_{\mathcal{T}_{\mathcal{L}}} \leq d^K \int d^{2N} \mathbf{x} f(\mathbf{x}) + \delta_{N,d} \quad (51)$$

for $f(\mathbf{x}) = e^{-\pi g |\mathbf{x}|^2}$ for $g > 0$ where $\delta_{N,d} \rightarrow 0$ as $d \rightarrow \infty$ and $d \ln(d) \ll N \ll e^{\pi g d}$. This result makes use of the balanced-ness of the set $\mathcal{T}$ containing all the self-orthogonal $[N, (N - K)/2]_{d^2}$ codes [25,67]. Additionally increasing d makes $\delta \rightarrow 0$ since the balanced-ness of this set allows us to relate the average over $\mathcal{T}_{\mathcal{L}}$ to a sum over the lattice $\mathbb{Z}^{2N}/\sqrt{d}$. Note that any concatenated lattice is always a sublattice of $\mathbb{Z}^{2N}/\sqrt{d}$ and evaluating this sum over this lattice lets us directly relate it to the integral as a Riemann sum and the function $f(\mathbf{x}) = e^{-\pi g |\mathbf{x}|^2}$ is particularly well-behaved for this to hold true. While these are not exact relations, we group all our errors into the term $\delta_{N,d}$ which we show approaches zero along this defined sequence of $N, d \rightarrow \infty$. Hence we now state our main result.

Theorem 1. There exists a sequence of qudit stabilizer codes $[[N, K]]_d$ for d being prime with $d = 3 \pmod{4}$, with increasing N, d such that

$$\log_2(d) \frac{K}{N} = \log_2 \left(\frac{\eta}{1 - \eta} \right) - \tilde{\epsilon} \quad (52)$$

and $\tilde{\epsilon}$ can be made arbitrarily small simultaneously while the infidelity of this sequence of codes using transpose recovery after pure-loss of transmittance η converges to zero as $N, d \rightarrow \infty$ with $d \ln(d) \ll N \ll e^{\pi d \frac{\eta}{1-\eta}}$.

We refer the reader to Appendix G for a complete proof of the above result. We offer some more intuition for this result by noting that the increasing d makes the lattice $\mathbb{Z}^{2N}/\sqrt{d}$ finer. Any particular $[[N, K]]_d$ code is essentially defining some choice of lattice $\mathcal{L}^\perp \subseteq \mathbb{Z}^{2N}/\sqrt{d}$ which clearly increases the available choice on increasing d . Further this means that beginning with a fine enough lattice $\mathbb{Z}^{2N}/\sqrt{d}$ lets us approximate a good lattice which is capable of good sphere packing. We motivate this idea in Fig. 6. Hence we find that through the concatenation of the appropriate stabilizer code with square GKP qudits, the capacity of the loss channel is achieved as a rate for all values of transmission. While our proof is restricted to only prime

values of d with $d = 3 \pmod{4}$, we expect this to be generalizable for any d which is a prime power and mainly use $d = 3 \pmod{4}$ since the basis choice for $GF(d^2)$ allows for a relatively clearer mapping to the lattice formalism.

VI. DISCUSSION

In this work we have shown that the rates of coherent information for the Gaussian displacement noise channel, capacity of the pure-loss channel and the amplification channel are all achievable through the use of concatenated GKP square qudits by taking the limit of very large individual qudit dimension. While this limit may seem unphysical, we highlight that for reasonably small $d = 5$ and $d = 7$, we can already get incredibly close to the coherent information of the Gaussian displacement noise channel. This is especially encouraging given recent experimental progress in the creation of square GKP qudits [69]. In addition for the displacement noise channel, we are able to explicitly construct sequences of these codes through the use of quantum polar codes. This result hinges on the fact that the effective logical noise after correction for the displacement noise channel can be reduced into a classical mixture of qudit Pauli noise with an auxiliary system containing syndrome information.

While our results have focused on infinite-energy GKP for the sake of finding achievable rates, we note that the method of using analog information can be extended to the finite energy case. Here we work under the twirling approximation used in previous works [33,36] to treat finite-energy GKP as a Gaussian displacement noise channel on an infinite-energy GKP. If the innate noise model is purely Gaussian displacement noise with spread σ_0 , making the data modes finite-energy just increases the effective noise to $\sqrt{\sigma_0^2 + \sigma_{\text{data}}^2}$ where $\sigma_{\text{data}}^2 = \tanh(\Delta^2/2)$. However, we note that if the ancillas are finite-energy, the syndrome information obtained in the ancillas are no longer faithful. This is since the ancillary modes itself have some random displacement noise associated to it. We can exactly quantify this uncertainty by accounting for it in a new probability distribution (assuming base-noise strength σ)

$$\begin{aligned} p_{\sigma_{\text{anc}}}(u, s_1) &= \frac{1}{2\sigma_{\text{anc}}} \int d\xi e^{-\xi^2/4\sigma_{\text{anc}}^2} p \left(u, \left(s_1 + \xi \frac{d}{2\pi} \right) \pmod{1} \right) \\ &= \frac{1}{d} \theta_3 \left(\pi \left(\frac{u + s_1}{d} \right), e^{-\pi(\sigma^2 + 2\sigma_{\text{anc}}^2)/d} \right) \end{aligned} \quad (53)$$

where we have considered a teleportation-based protocol for syndrome extraction [34] and $\theta_3(u, q) = \sum_{n \in \mathbb{Z}} q^{n^2} \cos(2nu)$ is a Jacobi-theta function. Note that this means that we can account for both finite-energy ancilla and data modes by increasing the effective noise

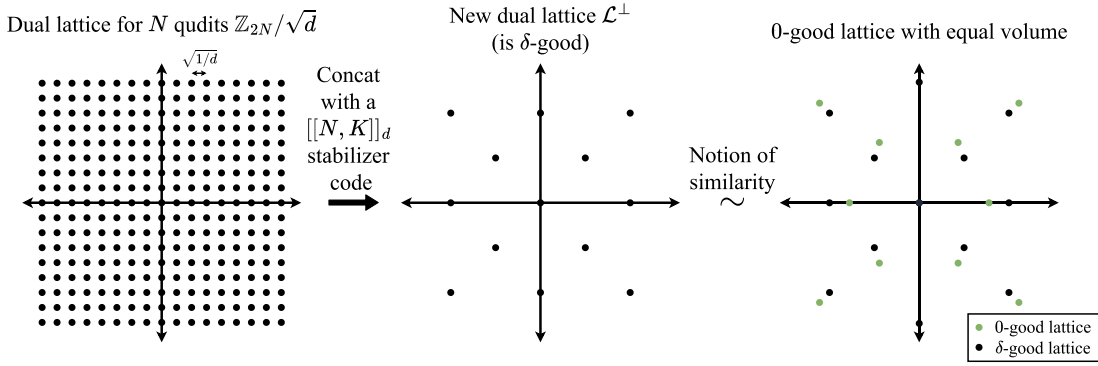


FIG. 6. A rough visual depiction of the working principle behind how qudit stabilizer codes concatenated to square GKP qudits can be used to achieve the capacity of the pure-loss channel. Considering N square GKP qudits by themselves, the dual lattice of this N mode GKP code corresponds to the integer lattice scaled down by $\sqrt{d}$ which is $\mathbb{Z}^{2N}/\sqrt{d}$. Concatenation with some outer $[[N, K]]_d$ code restricts what can be considered a valid logical operation hence choosing some sublattice $\mathcal{L}^\perp \subseteq \mathbb{Z}^{2N}/\sqrt{d}$. A crucial aspect of increasing d here shows that this allows for $\mathcal{L}^\perp$ to be chosen to have certain good properties. We note that this means that we can define some kind of notion of similarity between this concatenated lattice and another lattice of the same volume which is 0-good (which has direct capacity achieving properties). We base this notion of similarity using δ -good as a definition where as $\delta \rightarrow 0$, we are able to capture the capacity achieving properties a 0-good lattice would have. These lattices need not actually be similar in their exact descriptions but for the sake of illustration we depict them to be similar in the above figure. Using this δ -good property, we are able to claim the existence of asymptotically good lattices to show that the capacity of pure-loss and the amplification channel is an achievable rate.

$\sigma_{\text{eff}} = \sqrt{\sigma_0^2 + \sigma_{\text{data}}^2 + 2\sigma_{\text{anc}}^2}$. This shows a possible way of achieving the rate $I_{d,\text{analog}}^{\text{sq}}(\sigma_{\text{eff}})$ if the base noise strength is σ_0 .

However, quantum polar codes have high weight stabilizers which poses a direct problem to making this a practical implementation using such codes. There are ways of fault tolerantly extracting syndrome information for quantum polar codes [59] and one could also use a Knill-Glancy scheme for the overall GKP code obtained through concatenation [70] which would give $2N$ syndromes that are sufficient for decoding. However, these are resource heavy since they require preparation of an ancilla system which is encoded using the polar code, and would hence require a fault-tolerant encoding method for the outer polar code. Given the structure of the polar code encoder, we do not expect this to be an easy task on its own. Given the extensive amount of work done in constructing fault-tolerant methods of logical operations as well as syndrome extraction for QLDPC codes [71–76], we note that a lot of these directly can be extended with our results for GKP codes while still achieving non-zero rates since we have only examined concatenated square GKP codes. We note that QLDPC-GKP concatenated codes that use analog information has been explored for the square GKP qubit case [22,77].

In regard to our results in the case of the pure-loss and amplification channels we would like to note that it is currently limited to an existence-based result. Given that we find a capacity achieving sequence through qudit stabilizer codes, we believe that stronger claims on what

these stabilizer codes can be possibly made by restricting to polynomial codes such as non-binary Reed-Muller codes [78–80] and leave this exploration for future work. Importantly, this result adds more understanding to the structure of what can be considered a good lattice since we are able to express it as a concatenated square lattice. These results further highlight the fundamental connection between lattices and error correcting codes [67,81] and also offer an exploration of goodness for sphere-packing through our definition of a δ -good lattice. We hope that our results spark more interest in exploration in regards to concatenated GKP codes.

ACKNOWLEDGMENTS

We would like to thank Kyungjoo Noh, Filip Rozpędek, Gideon Lee, Debayan Bandopadhyay, Mariesa Teo, and Han Zheng for useful discussions. We acknowledge support from the ARO (W911NF-23-1-0077), ARO MURI (W911NF-21-1-0325), AFOSR MURI (FA9550-21-1-0209, FA9550-23-1-0338), DARPA (HR0011-24-9-0359, HR0011-24-9-0361), NSF (ERC-1941583, OMA-2137642, OSI-2326767, CCF-2312755, OSI-2426975), Packard Foundation (2020-71479), and the Marshall and Arlene Bennett Family Research Program. This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers and Advanced Scientific Computing Research (ASCR) program under contract number DE-AC02-06CH11357 as part of the InterQnet quantum networking project.

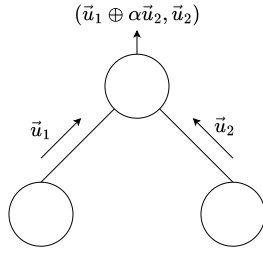


FIG. 7. A unit of the encoding operation. The two leaves here take equal length dit-vectors and then do a ditwise xor between $\vec{u}_1$ and $\alpha \vec{u}_2$ while retaining a copy of $\vec{u}_2$ to output a dit-vector of twice the length.

DATA AVAILABILITY

The data that support the findings of this article are openly available [82].

APPENDIX A: A PRIMER TO POLAR CODES

1. Encoding

The encoding scheme for polar codes [48] can be understood as moving up a binary tree of depth $\log N$ to encode N bits. The basic building block is shown in Fig. 7 with the whole encoding tree depicting the operation G_N in Fig. 8. This is then used to construct a binary tree for the whole process. The complexity of encoding is $\mathcal{O}(N \log N)$ operations since each layer of the binary tree has a total of $N/2$ operations and there are $\log N$ layers. Here all addition is modulo d where d refers to the dimension of the dit.

Definition A1 (Polar transform). The polar transform over $N = 2^n$ dits $P_N : \mathbb{F}_d^N \rightarrow \mathbb{F}_d^N$ is defined by the action of the matrix G_N on the N bit vector u_1^N as $P_N(u_1^N) = G_N u_1^N$ where all addition is modulo d and

$$G_N = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}^{\otimes n}. \quad (\text{A1})$$

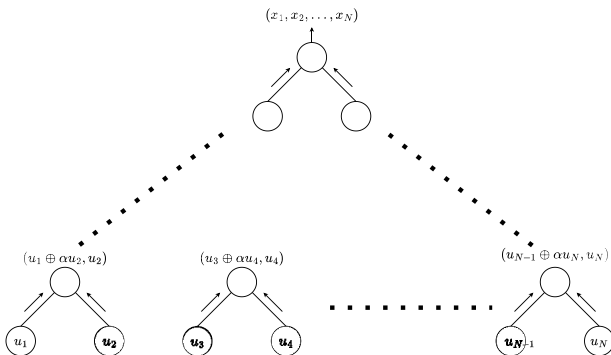


FIG. 8. The full encoding binary tree for N dits with the output $x_1^N = G_N u_1^N$.

The standard polar transform uses $\alpha = 1$, but any choice $\alpha \in \mathbb{F}_d \setminus \{0\}$ can be used. We will be referring to α as the kernel choice.

Definition A2 (Polar code $(N, K, \mathcal{A}, u_{\mathcal{A}_c})_{d, \alpha}$). A polar code $(N, K, \mathcal{A}, u_{\mathcal{A}_c})_{d, \alpha}$ is a code of block-length N which encodes messages carrying $K = |\mathcal{A}|$ dits of information where $\mathcal{A} \subseteq \{1, \dots, N\}$ and $u_{\mathcal{A}_c} \in \mathbb{F}_d^{N-K}$. The codewords for this code are obtained by encoding information in the bits with indices in set $\mathcal{A}$ and freezing all the other bits in $\mathcal{A}_c$ with the values $u_{\mathcal{A}_c}$ and then performing the polar transform with kernel α on these input bits. Alternatively if all $u_{\mathcal{A}_c} = 0$, we shorten this notation to $(N, K, \mathcal{A})_{d, \alpha}$.

2. Understanding channel polarization

In this section we provide a short introduction to the way channel recombining and splitting works and how it creates the conditions that lets channel polarization take place. We provide a definition of the discrete memoryless channel (DMC) as follows.

Definition A3 (Discrete memoryless channel). We define a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ where for any $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, $W(y|x)$ equals the probability of obtaining output y when the input was x . For any valid DMC, $\sum_{y \in \mathcal{Y}} W(y|x) = 1$. At any given time, the output of a DMC is only dependent on the input at that particular time, hence being memoryless.

We begin by considering the application of the kernel over two dits $u_1, u_2 \in \mathbb{F}_d$ where we take $(u_1, u_2) \rightarrow (u_1 + \alpha u_2, u_2)$ with $\alpha \in \mathbb{F}_d \setminus \{0\}$. Assuming that the output dits $x_1 = u_1 + \alpha u_2$ and $x_2 = u_2$ both experience noise channel $W : \mathbb{F}_d \rightarrow \mathcal{Y}$, we can consider the two outputs to be y_1 and y_2 .

We now consider the probabilities of obtaining the pair of outputs (y_1, y_2) from the input u_1 considering all u_2 to be possible following which we consider the probability of obtaining (y_1, y_2) from the input u_2 assuming a particular value for u_1 . This can be similarly thought of as a noise channel mapping u_1 to the output (y_1, y_2) and then mapping u_2 to the output (y_1, y_2, u_1) . This lets us describe the previous process of obtaining (y_1, y_2) from (x_1, x_2) into two separated channels that we label as W^+ and W^- described by transition probabilities

$$\begin{aligned} W^-(y_1, y_2 | u_1) &= \frac{1}{d} \sum_{u_2 \in \mathbb{F}_d} W(y_1 | u_1 + \alpha u_2) W(y_2 | u_2) \\ W^+(y_1, y_2, u_1 | u_2) &= \frac{1}{d} W(y_1 | u_1 + \alpha u_2) W(y_2 | u_2). \end{aligned} \quad (\text{A2})$$

A crucial feature of these newly obtained channels is that it lets us fully describe the noise process in terms of channels

that act on the same input alphabet $\mathcal{X} = \mathbb{F}_d$. Additionally, the channels W^+ and W^- are purely described in terms of the same channel W .

Suppose we consider a polar code of size N , and consider the noise process for the input dit u_i to the output (y_1^N, u_1^{i-1}) by the channel $W^{(i),N} : \mathcal{X} \rightarrow \mathcal{Y}^N \times \mathcal{X}^{i-1}$ which assumes the dits u_{i+1}^N to take on any possible value. This would be described by transition probabilities in Eq. (A3)

$$W_N^{(i)}((y_1^N, u_1^{i-1})|u_i) = \sum_{u_{i+1}^N \in \mathbb{F}_d^{N-i}} \frac{1}{d^{N-i}} \left[\prod_{j=1}^N W(y_j | (P_N(u_1^N))_j) \right]. \quad (\text{A3})$$

We will later see why this form of writing the channel makes describing the successive cancellation decoding algorithm very straightforward. From the description of encoding in Fig. 8, it is clear that if we now double the blocklength to $2N$, we can consider it to be composed of two separate polar code encoders for blocklength N that take input of $\vec{u}_1 + \alpha \vec{u}_2$ and $\vec{u}_2$, the outputs of which become the odd and even entries of the $2N$ length dit-string x_1^{2N} . This lets us apply the formula describing W^+ and W^- in terms of W to describe $W_{2N}^{(2i)}$ and $W_{2N}^{(2i-1)}$ completely in terms of $W_N^{(i)}$ as

$$W_{2N}^{(2i-1)}((y_1^{2N}, u_1^{2i-2})|u_{2i-1}) = \frac{1}{d} \sum_{u_{2i} \in \mathbb{F}_d} W_N^{(i)}((y_{1,o}^{2N}, u_{1,o}^{2i-2} + \alpha u_{2,e}^{2i-2})|u_{2i-1} + \alpha u_{2i}) \times W_N^{(i)}((y_{2,e}^{2N}, u_{2,e}^{2i-2})|u_{2i}), \quad (\text{A4})$$

$$W_{2N}^{(2i)}((y_1^{2N}, u_1^{2i-1})|u_{2i}) = \frac{1}{d} W_N^{(i)}((y_{1,o}^{2N}, u_{1,o}^{2i-2} + \alpha u_{2,e}^{2i-2})|u_{2i-1} + \alpha u_{2i}) \times W_N^{(i)}((y_{2,e}^{2N}, u_{2,e}^{2i-2})|u_{2i}), \quad (\text{A5})$$

where the notation $u_{1,o}^{2N}$ is the length N subset of u_1^{2N} composed of the terms with odd indices and $u_{2,e}^{2N}$ is the length N subset of u_1^{2N} composed of the terms with even indices.

We now introduce some more specific Bhattacharya distances for telling apart distributions $W(y|x)$ and $W(y|x')$ as

$$Z(W_{\{x,x'\}}) = \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x')} \quad (\text{A6})$$

following which we define an averaged version of the above expression which fixes the difference between x and

x' to be

$$Z_\Delta(W) = \frac{1}{d} \sum_{x \in \mathcal{X}} Z(W_{\{x,x+\Delta\}}). \quad (\text{A7})$$

For any channel W , the one shot maximum likelihood error satisfies [46]

$$P_e \leq \sum_{\Delta \in \mathbb{F}_d \setminus \{0\}} Z_\Delta(W) \quad (\text{A8})$$

which is relevant to us in the way we have defined the channels $W_N^{(i)}$. As we will describe in the following subsection, the successive cancellation decoder is performing a one-shot maximum likelihood decoding for each dit u_i for the noise channel $W_N^{(i)}$. Importantly this is not the same as an overall maximum likelihood decoder since this does not take into account the fact that for all $i \in \mathcal{A}_c$, the value of u_i is frozen.

Using similar calculations to that of Ref. [46], we have

$$Z_\Delta(W^+) = Z_{\alpha u}(W)Z_u(W) \quad (\text{A9})$$

$$Z_\Delta(W^-) \leq 2Z_\Delta(W) + \sum_{\Delta' \in \mathbb{F}_d \setminus \{0, -\alpha^{-1}\Delta\}} Z_{\Delta'}(W)Z_{\Delta+\alpha\Delta'}(W). \quad (\text{A10})$$

Since each of the channels $W_N^{(i)}$ can be understood as a sequence of $W^{b_0 b_1 \dots b_n}$ where $N = 2^n$ and $b_i \in \{+, -\}$, we can use the above inequality to actually define upper bounds for each of the values of $Z_\Delta(W_{2N}^{(i)})$ with respect to the values of $Z_\Delta(W_N^{(j)})$.

Treating the sequence $b_0 b_1 \dots b_n$ as a randomized sequence where each b_i is equally likely to be $+$ or $-$, $I(W^{b_0 b_1 \dots b_n})$ becomes a martingale sequence using the inequalities in Eqs. (A10) and (32) as is highlighted in Ref. [46] which converges either to 0 or 1. The main requirements for this proof of polarization is that for any channel $W^{b_0 b_1 \dots b_i}$ along this sequence, the new channels obtained through the recombination of $W^{b_0 b_1 \dots b_i}$ which would be $W^{b_0 b_1 \dots b_i -}$ and $W^{b_0 b_1 \dots b_i +}$ must satisfy $Z(W^{b_0 b_1 \dots b_i +}) \leq Z(W^{b_0 b_1 \dots b_i})^2$ and $Z(W^{b_0 b_1 \dots b_i -}) \leq qZ(W^{b_0 b_1 \dots b_i})$ for some $q > 0$. This can be seen to trivially follow from Eq. (A10) along with the condition that $Z(W) \rightarrow 0$ iff $I(W) \rightarrow 1$ and $Z(W) \rightarrow 1$ iff $I(W) \rightarrow 0$ for any channel W . We refer the reader to Refs. [47,50,54,83] for further studies of the working principle of polarization. As such we now have the sufficient conditions to have $Z(W_N^{(i)}) \leq 2^{-N^{-\beta}}$ for $0 < \beta < 1/2$ hold for K indices i with $K/N \rightarrow I(W)$ in the limit of large enough N (see Lemma 2 in Ref. [46]). This lets us guarantee that there would be a capacity achieving sequence of polar codes that can in theory be explicitly found which has the error probabilities upper bounded by the sequence $N2^{-N^{-\beta}}$ for a particular noise channel W .

3. Successive cancellation decoding

Here we describe the decoding procedure when each dit of the codewords experience the same DMC noise W . The decoding procedure makes use of the binary tree structure to achieve a complexity of $N \log(N)$ by making a decision of each bit only using the information of the output and the decisions of the bits which precede it.

To illustrate this idea, we first examine a single unit of the decoding procedure. Here we have $x_1 = u_1 \oplus \alpha u_2$ and $x_2 = u_2$. We have given with us the following log-likelihood ratios:

$$\mathbf{L}_j = L_j^{[i]} = \log\left(\frac{p(x_j = 0)}{p(x_j = i)}\right), \quad i \in \{0, \dots, d-1\}. \quad (\text{A11})$$

Here $p(\text{event})$ simply refers to the overall probability of said event occurring under a certain fixed circumstance which in this case would be corresponding to receiving noisy outputs y_1 and y_2 which correspond to inputs of x_1 and x_2 through noise channel W . We now wish to know the log-likelihood ratios for the bit u_1 agnostic to what the value of u_2 is. It then follows that

$$p(u_1 = i) = \sum_{j=0}^{d-1} p(x_1 = i \oplus \alpha j) p(x_2 = j), \quad (\text{A12})$$

which then gives us

$$\log\left(\frac{p(u_1 = 0)}{p(u_1 = i)}\right) = f_\alpha(\mathbf{L}_1, \mathbf{L}_2)^{[i]}, \quad (\text{A13})$$

where we define the function $f_\alpha(\mathbf{L}_1, \mathbf{L}_2)$ taking two d length vectors of log-likelihood ratios (LLRs) and outputs one d length vector of LLRs as

$$\begin{aligned} f_\alpha(\mathbf{L}_1, \mathbf{L}_2)^{[i]} &= \log\left(\sum_{j=0}^{d-1} \exp\left(-(L_1^{[\alpha j]} + L_2^{[j]})\right)\right) \\ &\quad - \log\left(\sum_{j=0}^{d-1} \exp\left(-(L_1^{[\alpha j \oplus i]} + L_2^{[j]})\right)\right) \end{aligned} \quad (\text{A14})$$

which we further generalize to taking inputs of two lists of M different d length vectors of LLRS defined as $\mathbf{L}_1^M$ and $\mathbf{K}_1^M$ respectively to be

$$\begin{aligned} f_\alpha(\mathbf{L}_1^M, \mathbf{K}_1^M)_m^{[i]} &= \log\left(\sum_{j=0}^{d-1} \exp\left(-(L_m^{[\alpha j]} + K_m^{[j]})\right)\right) \\ &\quad - \log\left(\sum_{j=0}^{d-1} \exp\left(-(L_m^{[\alpha j \oplus i]} + K_m^{[j]})\right)\right), \end{aligned} \quad (\text{A15})$$

where $f_\alpha(\mathbf{L}_1^M, \mathbf{K}_1^M)$ is also a list of M different d length vectors and the index m in the above expression goes from 1 to M .

If we assume we have the exact knowledge of what u_1 is, we can use it to determine the probabilities for u_2 once this knowledge is included in the event

$$p(u_2 = i | u_1) = p(x_1 = u_1 \oplus \alpha i) p(x_2 = i). \quad (\text{A16})$$

In terms of the LLRs, we have,

$$\log\left(\frac{p(u_2 = 0 | u_1)}{p(u_2 = i | u_1)}\right) = g_\alpha(\mathbf{L}_1, \mathbf{L}_2, u_1)^{[i]}, \quad (\text{A17})$$

where we define the function $g_\alpha(\mathbf{L}_1, \mathbf{L}_2, u_1)$ taking two d length vectors of LLRs and one dit (u) and outputs one d length vector of LLRs as

$$g_\alpha(\mathbf{L}_1, \mathbf{L}_2, u)^{[i]} = L_2^{[i]} - L_1^{[u]} + L_1^{[u \oplus \alpha i]}, \quad (\text{A18})$$

which we also further to taking inputs of two lists of M different d length vectors of LLRS defined as $\mathbf{L}_1^M$ and $\mathbf{K}_1^M$ respectively along with a list of M dits u_1^M as

$$g_\alpha(\mathbf{L}_1^M, \mathbf{K}_1^M, u_1^M)_m^{[i]} = K_m^{[i]} - L_m^{[u_m]} + L_m^{[u_m \oplus \alpha i]}. \quad (\text{A19})$$

Once we have the LLRs for any particular bit $\mathbf{L}$, we make the guess of what the dit value is based on the function

$$\text{dit}(\mathbf{L}) = \begin{cases} 0 & \min_i L^{[i]} \geq 0 \\ \text{argmin}_i L^{[i]} & \text{otherwise} \end{cases}. \quad (\text{A20})$$

The above function selects the most likely value for the dit. Assuming we have knowledge of the outputs and the nature of the noise channel, we can calculate the input LLRs vector as $\mathbf{L}_1^N$ where

$$L_i^{[j]} = \log\left(\frac{W(y_j | 0)}{W(y_j | j)}\right), \quad i \in \{1, \dots, N\}, j \in \{0, \dots, d-1\}. \quad (\text{A21})$$

The beliefs of the left child of the root in the encoding tree can be calculated to simply be $f_\alpha(\mathbf{L}_i, \mathbf{L}_{i+N/2})$ where $i = 1$ to $N/2$. Suppose we now have a successive cancellation decoder for $N/2$ bits, we simply use the $N/2$ values of $f_\alpha(\mathbf{L}_i, \mathbf{L}_{i+N/2})$ as input to those to get guesses $\hat{u}_1^{N/2}$ which is the guessed output of the left child to the root node. We now can pass beliefs $g_\alpha(\mathbf{L}_i, \mathbf{L}_{i+N/2}, \hat{u}_i)$ where $i = 1$ to $N/2$ to the right child of the root and use the SC decoder for $N/2$ bits with these inputs and get a guess of what the outputs of the right child must have been which will finally give us the guess of the decoded codeword $\hat{x}_1^N$. The recursive nature of this is captured by the steps L , R and then U in Fig. 9. The channel which dit u_i experiences in this can be

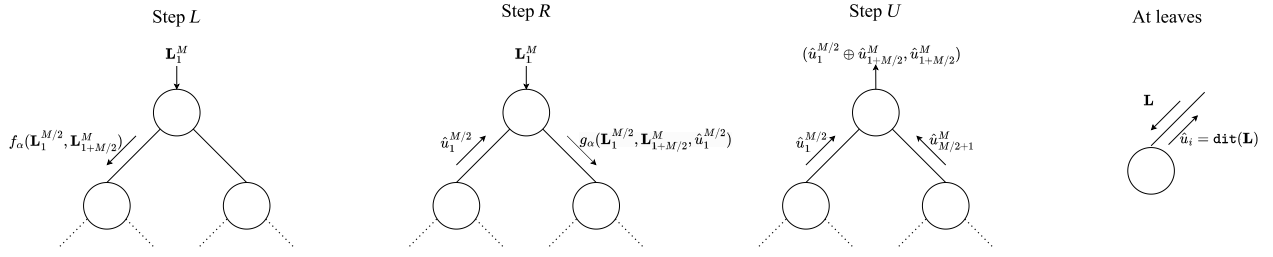


FIG. 9. A unit for the decoding steps at each node. The steps proceed as step L (pass values to left child), step R (pass values to right child) and then step U (pass guessed dits to parent). If it reaches a leaf, the leaf returns a guess $\hat{u}_i$ using the function dit .

written as $W^{(i)}((y_1^N, u_1^{i-1})|u_i)$ described in Eq. (A3) and the beliefs that reach the leaf nodes in the tree are simply the LLRs associated to u_i with this channel.

Theorem A1 (Classical polar coding). For a discrete memoryless channel W with input alphabet $\mathcal{X}$ with cardinality d that is prime and an output alphabet $\mathcal{Y}$, there exists a sequence of polar code under successive cancellation decoding which can achieve coding rate K/N arbitrarily close to

$$I(W) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{d} W(y|x) \log_d \left(\frac{W(y|x)}{\sum_{x' \in \mathcal{X}} \frac{1}{d} W(y|x')} \right). \quad (\text{A22})$$

Here $W(y|x)$ is the probability of obtaining y as a noisy output for input x .

Proof. See Ref. [46]. ■

Note that while the analysis in Ref. [46] is done for discrete memoryless channels with a finite size of output alphabet, all their analysis can be extended to channels with a continuous output alphabet.

APPENDIX B: POLAR CODES FOR CLASSICAL MIXTURES OF PAULI CHANNELS

Definition B1 (Quantum polar transform). We define the polar transform in the qudit Z basis as follows:

$$\hat{V} = \sum_{\mathbf{z} \in \mathbb{F}_d^N} |G_N \mathbf{z}\rangle \langle \mathbf{z}|. \quad (\text{B1})$$

Where all operations are performed modulo d to stay in the field of $\mathbb{F}_d^N$. Using the standard definitions of the qudit X and Z bases being defined as the bases which diagonalize

the following operators:

$$\hat{X} = \sum_{j=0}^{d-1} |j+1\rangle \langle j|, \quad \hat{Z} = \sum_{j=0}^{d-1} \omega^j |j\rangle \langle j|, \quad (\text{B2})$$

we can rewrite the quantum polar transform in the qudit x basis states $|\tilde{\mathbf{x}}\rangle$ by noting

$$\begin{aligned} V &= \sum_{\mathbf{z} \in \mathbb{F}_d^N} |G_N \mathbf{z}\rangle \langle \mathbf{z}| \\ &= \sum_{\mathbf{z}, \mathbf{x}, \mathbf{x}' \in \mathbb{F}_d^N} \frac{e^{\frac{2\pi i}{d} (\mathbf{x}' \cdot G_N \mathbf{z} - \mathbf{x} \cdot \mathbf{z})}}{d} |\tilde{\mathbf{x}}'\rangle \langle \tilde{\mathbf{x}}| \\ &= \sum_{\mathbf{x} \in \mathbb{F}_d^N} |G_N^{-T} \tilde{\mathbf{x}}\rangle \langle \tilde{\mathbf{x}}|. \end{aligned} \quad (\text{B3})$$

In the last step, we have used the property that $\frac{e^{\frac{2\pi i}{d} (\mathbf{x}' \cdot G_N \mathbf{z} - \mathbf{x} \cdot \mathbf{z})}}{d}$ will sum to 1 if and only if $\mathbf{x}' \cdot G_N \mathbf{z} - \mathbf{x} \cdot \mathbf{z} = 0 \pmod{d}$ otherwise it sums to zero. Hence giving the condition $G_N^T \mathbf{x}' = \mathbf{x}$. This is in effect the same as saying that all CSUM gates in the $\mathbf{z}$ basis become into CSUM⁻¹ gates in the $\mathbf{x}$ basis with the control and target swapped. With the above definition, we can note that the polar transform acts simultaneously in the X and Z bases with the order of qudits reversed from one to the other and a different kernel since

$$G_N^{-T} = \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}^{\otimes n}. \quad (\text{B4})$$

However, as is pointed out in Ref. [54], this kernel also polarizes the channels in the same way.

Theorem B1 (Quantum polar code achievable rate for Pauli noise). For a qudit (d levels which is prime) Pauli noise described by

$$\mathcal{N}(\cdot) = \sum_{u,v} p_{u,v} \hat{X}^u \hat{Z}^v (\cdot) (\hat{X}^u \hat{Z}^v)^\dagger, \quad (\text{B5})$$

there exists a sequence of quantum polar codes using the SC decoder that asymptotically achieve the rate (in qudit

per channel use)

$$R = I(W_A) + I(W_P) - 1, \quad (\text{B6})$$

where

$$W_A(z + u|z) = \sum_v p_{u,v} = p_u, \quad W_P((x + v, u)|x) = p_{u,v}. \quad (\text{B7})$$

Proof. We define for our convenience the following channels:

$$W_A^c(z + u|z) = W_A(z + d - u|z), \quad (\text{B8})$$

$$W_P^c((x + v, u)|x) = W_P((x + d - v, u)|x).$$

Note that these channels are DMCs, and so will polarize for prime d , and represent a noise channel that acts on the amplitude (z) and phase (x) bases respectively hence some of the qudits will be good for amplitude and some will be good for phase. Let us consider two parties A and B . A has N qudit Bell pairs and sends N of these halves through a noise channel to B . We can divide N into four sets based on the channels W_A^c and W_P^c .

- (1) $\mathcal{I}$ is the set of qudits which are good for both amplitude and phase.
- (2) $\mathcal{A}$ is the set of qudits bad for amplitude but good for phase.
- (3) $\mathcal{P}$ is the set of qudits bad for phase but good for amplitude.
- (4) $\mathcal{E}$ is the set of qudits bad for both amplitude and phase.

We can define projectors Π_A which freezes amplitude values for set $\mathcal{A}$ and Π_P which freezes phase values for set $\mathcal{P}$. If B decides to apply the polar transform on their qudits while the frozen values are decided beforehand by A and B , we end up in the state

$$|\psi_0\rangle = \frac{1}{\sqrt{d^{N-|\mathcal{A} \cup \mathcal{P}|}}} \sum_z \Pi_A \Pi_P |z\rangle_A \otimes |G_N z\rangle_B. \quad (\text{B9})$$

Now B suffers noise on their qudits (perhaps after some communication channel) which is described by the action of $\mathcal{N}$ on each qudit. We now consider an additional environment mode which tracks the error that has occurred hence ending up in the state

$$|\psi_1\rangle \propto \sum_{z,u,v} \sqrt{p_{u,v}} \Pi_A \Pi_P |z\rangle_A \otimes X^u Z^v |G_N z\rangle_B \otimes |\mathbf{u}, \mathbf{v}\rangle_E. \quad (\text{B10})$$

Applying the inverse of the polar transform defined by

$$V^\dagger = \sum_z |G_N^{-1} z\rangle \langle z| = \sum_x |G_N^T \tilde{\mathbf{x}}\rangle \langle \tilde{\mathbf{x}}|, \quad (\text{B11})$$

on B gives

$$|\psi_2\rangle \propto \sum_{z,u,v} \sqrt{p_{u,v}} \Pi_A \Pi_P |z\rangle_A \otimes \hat{X}^u \hat{Z}^v |z\rangle_B \otimes |\mathbf{u}, \mathbf{v}\rangle_E. \quad (\text{B12})$$

We note the following

$$\hat{V}^\dagger \hat{X}^u \hat{V} = \hat{X}^{u'}, \quad G_N u' = u, \quad (\text{B13})$$

$$\hat{V}^\dagger \hat{Z}^v \hat{V} = \hat{Z}^{v'}, \quad G_N^{-T} v' = v, \quad (\text{B14})$$

which highlights how $\mathbf{u}$ is obtained by the polar transform G_N on $\mathbf{u}'$ and similarly $\mathbf{v}$ is obtained by the polar transform G_N^{-T} on $\mathbf{v}'$. By measuring qudits in $\mathcal{A} \cup \mathcal{E}$ in the amplitude basis, we can obtain the dit values of $\mathbf{u}'$ corresponding to those. Further we know that $\mathbf{u}$ is nothing but the noisy output of the input $\mathbf{0}$ to the channel W_A which is equivalent to treating $\mathbf{0}$ as the noisy output to the input $\mathbf{u}$ to the channel W_A^c since $W_A^c(0|u) = W_A(u|0)$. Since $\mathbf{u} = G_N \mathbf{u}'$, it is a codeword for the n -dit classical polar code with dits $|\mathcal{A} \cup \mathcal{E}|$ frozen to the values obtained on the measurement of those qudits in the amplitude basis. Hence the task of finding $\mathbf{u}'$ is simply the decoding of the classical polar code with noisy output of $\mathbf{0}$ from the channel W_A^c which from Theorem A1 will achieve a rate of $I(W_A^c) - \epsilon_1$ for $\epsilon_1 > 0$. Note that trivially $I(W_A^c) = I(W_A)$ and similarly $I(W_P^c) = I(W_P)$ since they only differ by a permutation.

After this decoding is complete we obtain some estimate of the amplitude error $\hat{\mathbf{u}}$ which on correcting, we end up in the state

$$|\psi_3\rangle \propto \sum_{x,v} \sqrt{p_{v|\hat{\mathbf{u}}}} \Pi_A \Pi_P |\tilde{\mathbf{x}}\rangle_A \otimes Z^{v'} |\tilde{\mathbf{x}}\rangle_B \otimes |\mathbf{0}, \mathbf{v}\rangle_E + \mathcal{O}(\sqrt{p_A}), \quad (\text{B15})$$

where $\sqrt{p_A}$ is the probability that the classical decoder for noise channel W_A^c fails and as a result will contribute a small mixture of errors in the output state as a $\mathcal{O}(\sqrt{p_A})$ term. Similarly on measuring the qudits in set $\mathcal{P} \cup \mathcal{E}$ in the phase basis (note that we are able to do simultaneous phase basis and amplitude basis measurements for $\mathcal{E}$ since it consists of shared Bell pairs) we obtain the dit values of $\mathbf{v}'$ corresponding to those. Similarly we can note that $(\mathbf{0}, \hat{\mathbf{u}})$ is the noisy output of $\mathbf{v}$ (assuming the decoding of $\mathbf{u}$ is successful) through the noise channel W_P^c . This allows us to use the classical decoder again and this polar code achieves the rate of $I(W_P) - \epsilon_2$ for $\epsilon_2 > 0$. On decoding we will end up in the state

$$|\psi_3\rangle \propto \sum_{x,v} \Pi_A \Pi_P |z\rangle_A \otimes |z\rangle_B \otimes |\hat{\mathbf{u}}, \hat{\mathbf{v}}\rangle_E + \mathcal{O}(\sqrt{p_A}) + \mathcal{O}(\sqrt{p_P}). \quad (\text{B16})$$

We can make p_A and p_P arbitrarily small for any $\epsilon_1, \epsilon_2 > 0$.

Since all the sets $\mathcal{I}$, $\mathcal{A}$, $\mathcal{P}$ and $\mathcal{E}$ are disjoint, we have the following hold:

$$|\mathcal{A}| + |\mathcal{E}| = n(1 - I(W_A) + \epsilon_1), \quad (\text{B17})$$

$$|\mathcal{P}| + |\mathcal{E}| = n(1 - I(W_P) + \epsilon_2), \quad (\text{B18})$$

$$|\mathcal{I}| + |\mathcal{A}| + |\mathcal{P}| + |\mathcal{E}| = n. \quad (\text{B19})$$

We can encode information in set $\mathcal{I}$ but also there are pre-shared entangled pairs which are $\mathcal{E}$. Hence the effective rate is given by

$$R = \frac{|\mathcal{I}| - |\mathcal{E}|}{n} = I(W_A) + I(W_P) - 1 - (\epsilon_1 + \epsilon_2), \quad (\text{B20})$$

where ϵ_1 and ϵ_2 can be made arbitrarily small as n approaches infinity hence giving an asymptotic rate of $I(W_A) + I(W_P) - 1$. ■

Corollary B1. The quantum polar code achieves the Hashing bound of a Pauli noise channel.

Proof. The hashing bound for Pauli noise is defined as $1 - H(\mathbf{p})$ where $\mathbf{p}$ is the probability vector associated to the Pauli noise $p_{u,v}$. We can note the following:

$$I(W_P) = \sum_{u,v} p_{u,v} \log_d \left(\frac{p_{u,v}}{\frac{p_u}{d}} \right) = 1 - H(p_{u,v}) + H(p_u), \quad (\text{B21})$$

$$I(W_A) = \sum_u p_u \log_d \left(\frac{p_u}{\frac{1}{d}} \right) = 1 - H(p_u). \quad (\text{B22})$$

Hence giving

$$I(W_P) + I(W_A) - 1 = 1 - H(p_{u,v}) = 1 - H(\mathbf{p}), \quad (\text{B23})$$

which shows that quantum polar codes achieve the Hashing bound. ■

Now we can further generalize this idea to show that we can achieve the coherent information of a classical mixture of single qudit Pauli channels.

Theorem B2 (Quantum polar code achievable rate for CMP). For a qudit (d levels which is prime) Pauli noise described by

$$\mathcal{N}(\cdot) = \sum_{u,v} p_{u,v,s} \hat{X}^u \hat{Z}^v (\cdot) (\hat{X}^u \hat{Z}^v)^\dagger \otimes |\mathbf{s}\rangle \langle \mathbf{s}| \quad (\text{B24})$$

there exists a sequence of quantum polar codes using the SC decoder that asymptotically achieve the rate (in per

qudit)

$$R = I(W_A) + I(W_P) - 1 \quad (\text{B25})$$

where

$$\begin{aligned} W_A((z + u, \mathbf{s})|z) &= \sum_v p_{u,v,s} = p_u, \\ W_P((x + v, u, \mathbf{s})|x) &= p_{u,v,s} \end{aligned} \quad (\text{B26})$$

Proof. It is sufficient to show that there are classical polar codes which can achieve the rate $I(W_A)$ and $I(W_P)$ respectively. Note that these channels are simply memoryless channels with a finite sized input alphabet which is prime. It is known that the symmetric coherent information is achievable for these classical channels from Ref. [46]. Hence the rate R can be achieved by constructing a CSS code with entanglement assistance. ■

We believe that this can also be generalized to classical mixtures of more general multi-qudit Pauli noise, but would require a different kind of polar transforming kernel based on the Clifford channel combining principle introduced in Ref. [57] and leave proving this for future work.

APPENDIX C: LATTICE THETA FUNCTIONS

In this Appendix we list a few results related to lattice theta functions and introduce relevant notation for the same. We define the following Jacobi-theta functions:

$$\theta_1(u, q) = 2q^{1/4} \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)} \sin((2n+1)u) \quad (\text{C1})$$

$$\theta_2(u, q) = 2q^{1/4} \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)} \cos((2n+1)u) \quad (\text{C2})$$

$$\theta_3(u, q) = 1 + 2 \sum_{n=1}^{\infty} q^{n^2} \cos(2nu) \quad (\text{C3})$$

which we will be making use of in the following appendices. When $\theta_3(u, q)$ is evaluated at $u = 0$, we represent it as $\theta_3(q) = \theta_3(0, q)$. we define the lattice theta function for a lattice $\mathcal{L} \subseteq \mathbb{R}^{2N}$ as

$$\Theta_{\mathcal{L}}(q) = \sum_{\mathbf{x} \in \mathcal{L}} q^{|\mathbf{x}|^2}, \quad (\text{C4})$$

from which it follows that $\Theta_{\mathbb{Z}^N}(q) = \theta_3(q)^N$. The theta function of the euclidean dual of $\mathcal{L}$, defined as $\mathcal{L}^* = \{\mathbf{y} \in \mathbb{R}^{2N} | \mathbf{y} \cdot \mathbf{x} \in \mathbb{Z}\}$ is given by

$$\Theta_{\mathcal{L}^*}(e^{i\pi z}) = \det(\mathcal{L})(i/z)^N \Theta_{\mathcal{L}}(e^{-i\pi/z}), \quad (\text{C5})$$

where $z \in \mathbb{C}$. This relation is a special case of the Poisson summation formula shown in Ref. [84] and was also

noted in Ref. [16]. We consider $\mathcal{L}$ to be a symplectically integral lattice with the corresponding symplectic dual $\mathcal{L}^\perp$ [as defined in Eq. (D9)]. As noted in Ref. [16], $\mathcal{L}^\perp$ is related to $\mathcal{L}^*$ by a rotational transform and hence they have the same lattice theta functions $\Theta_{\mathcal{L}^\perp}(q) = \Theta_{\mathcal{L}^*}(q)$. We note that the relation in Eq. (C5) also allows us to write

$$\theta_3(e^{i\pi z}) = (i/z)^{1/2} \theta_3(e^{-i\pi/z}). \quad (\text{C6})$$

APPENDIX D: PAULI NOISE FROM CLOSEST LATTICE POINT GKP DECODING ON DISPLACEMENT NOISE

Here we adapt the notation introduced in Ref. [16]. For convenience of the reader, we list some of the preliminary notation used in Ref. [16], which we will also use. We have N bosonic modes with quadratures $\hat{q}$ and $\hat{p}$ using which we define $\hat{\mathbf{x}} = \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix}$ to represent a $2n$ -dimensional phase space. The commutation relations can be written as

$$[\hat{x}_k, \hat{x}_l] = iJ_{k,l}, \quad J = \begin{pmatrix} 0 & \mathbb{I}_n \\ -\mathbb{I}_n & 0 \end{pmatrix}. \quad (\text{D1})$$

The matrix J is the symplectic form and so we can define a displacement operator for a phase-space displacement $\boldsymbol{\xi} \in \mathbb{R}^{2n}$ as

$$\hat{D}(\boldsymbol{\xi}) = \exp \left\{ -i\sqrt{2\pi} \boldsymbol{\xi}^T J \hat{\mathbf{x}} \right\}, \quad (\text{D2})$$

which has the following properties:

$$\hat{D}^\dagger(\boldsymbol{\xi}) \hat{\mathbf{x}} \hat{D}(\boldsymbol{\xi}) = \hat{\mathbf{x}} + \sqrt{2\pi} \boldsymbol{\xi}, \quad (\text{D3})$$

$$\hat{D}(\boldsymbol{\xi}_1) \hat{D}(\boldsymbol{\xi}_2) = e^{-2\pi \boldsymbol{\xi}_1^T J \boldsymbol{\xi}_2} \hat{D}(\boldsymbol{\xi}_2) \hat{D}(\boldsymbol{\xi}_1), \quad (\text{D4})$$

$$\hat{D}(\boldsymbol{\xi}_1) \hat{D}(\boldsymbol{\xi}_2) = e^{-\pi \boldsymbol{\xi}_1^T J \boldsymbol{\xi}_2} \hat{D}(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2). \quad (\text{D5})$$

For a GKP lattice over n modes, we need exactly $2n$ displacements to generate the stabilizer group given by

$$\mathcal{S} = \langle \hat{D}(\boldsymbol{\xi}_1), \dots, \hat{D}(\boldsymbol{\xi}_{2n}) \rangle. \quad (\text{D6})$$

These displacements must be linearly independent and must commute with each other, giving the condition $\boldsymbol{\xi}_j^T J \boldsymbol{\xi}_k \in \mathbb{Z}$ for all $j, k \in \{1, \dots, 2n\}$. We now define the matrix

$$M = \begin{pmatrix} \boldsymbol{\xi}_1^T \\ \boldsymbol{\xi}_2^T \\ \vdots \\ \boldsymbol{\xi}_{2n}^T \end{pmatrix}, \quad (\text{D7})$$

using which we define the symplectically integral lattice

$$\mathcal{L} = \{ \boldsymbol{\xi} \in \mathbb{R}^{2n} | \boldsymbol{\xi} = M^T \mathbf{a}, \mathbf{a} \in \mathbb{Z}^{2n} \}. \quad (\text{D8})$$

A GKP code defined using the following lattice would have its logical operations corresponding to displacements

that are present in the dual lattice which would be defined by

$$\mathcal{L}^\perp = \{ \boldsymbol{\xi}^\perp \in \mathbb{R}^{2n} | (\boldsymbol{\xi}^\perp)^T J \boldsymbol{\xi} \in \mathbb{Z}, \forall \boldsymbol{\xi} \in \mathcal{L} \}. \quad (\text{D9})$$

All the lattice points in $\mathcal{L}^\perp$ must necessarily be of form $\boldsymbol{\xi}^\perp = (MJ)^{-1} \mathbf{b}$ where $\mathbf{b} \in \mathbb{Z}^{2n}$. The lattice $\mathcal{L}$ is self-dual which is to say that $\mathcal{L} \subseteq \mathcal{L}^\perp$.

Since each stabilizer is a displacement, one can measure the eigenvalues of these displacements to obtain stabilizers. If we consider a state $|\psi\rangle$ which is stabilized by $\mathcal{S}$, we have

$$D(\boldsymbol{\xi}_i) D(\mathbf{e}) |\psi\rangle = e^{2\pi \boldsymbol{\xi}_i^T J \mathbf{e}} D(\mathbf{e}) |\psi\rangle. \quad (\text{D10})$$

which means we can measure the values $s_i = \boldsymbol{\xi}_i^T J \mathbf{e}$ modulo 1 without destroying the codeword information. These will be our syndrome measurements $\mathbf{s}$ each being modulo 1 and in the interval $[-\frac{1}{2}, \frac{1}{2})$ given by

$$\mathbf{s} = (MJ\mathbf{e}) \mod 1 = (MJ\mathbf{e}) - \lfloor MJ\mathbf{e} \rfloor, \quad (\text{D11})$$

for a particular value of $\mathbf{e}$ which we wish to correct. The initial guess for the correction is given by

$$\boldsymbol{\eta} = (MJ)^{-1} \mathbf{s}, \quad (\text{D12})$$

which would not account for possible logical errors. Using the closest-point decoding, one would find the closest lattice point in the dual lattice to this guessed correction.

This gives the effective correction to be

$$\bar{\boldsymbol{\eta}} = \boldsymbol{\eta} - \underset{\boldsymbol{\xi}^\perp \in \mathcal{L}^\perp}{\operatorname{argmin}} \|\boldsymbol{\eta} - \boldsymbol{\xi}^\perp\|, \quad (\text{D13})$$

for which the corresponding closest lattice point would be given by

$$\mathbf{b}_0(\mathbf{s}) = \underset{\mathbf{b} \in \mathbb{Z}^{2n}}{\operatorname{argmin}} (\mathbf{s} - \mathbf{b})^T (MM^T)^{-1} (\mathbf{s} - \mathbf{b}). \quad (\text{D14})$$

The logical error suffered by the correction of $\bar{\boldsymbol{\eta}}$ would be

$$\mathbf{e} - \bar{\boldsymbol{\eta}} = (MJ)^{-1} (\lfloor MJ\mathbf{e} \rfloor + \mathbf{b}_0). \quad (\text{D15})$$

Note that $\mathbf{e} - \bar{\boldsymbol{\eta}} \in \mathcal{L}^\perp$ hence is necessarily a Pauli error. Note that we can obtain the Pauli group for this particular multi-mode GKP code as $\mathcal{P} = \mathcal{L}^\perp / \mathcal{L}$ and for each particular $\mathbf{p} \in \mathcal{P}$, we can note that the probability of having occurred given that the syndrome is $\mathbf{s}$ would correspond to the occurrence of the event that $(MJ)^{-1} (\lfloor MJ\mathbf{e} \rfloor + \mathbf{b}_0) =$

$\mathbf{p} + \mathcal{L}$. This corresponds to having $\mathbf{e}$ be in the lattice defined by

$$\mathcal{L}_{\mathbf{p},\mathbf{s}} = \mathbf{p} + (MJ)^{-1}(\mathbf{s} - \mathbf{b}_0) + \mathcal{L} \quad (\text{D16})$$

which similarly can be extended to the lattice corresponding to a particular syndrome output as

$$\mathcal{L}_{\mathbf{s}} = (MJ)^{-1}(\mathbf{s} - \mathbf{b}_0) + \mathcal{L}^\perp. \quad (\text{D17})$$

Let us assume an underlying displacement noise channel of

$$\mathcal{N}(\cdot) = \int d^2\alpha P(\alpha) \hat{D}(\alpha)(\cdot) \hat{D}(\alpha)^\dagger. \quad (\text{D18})$$

We can now define the probability of doing a logical $\mathbf{p}$ Pauli operation and getting a syndrome of $\mathbf{s}$ corresponds to

$$p(\mathbf{p}, \mathbf{s}) \propto \sum_{\xi \in \mathcal{L}_{\mathbf{p},\mathbf{s}}} P(\xi) \quad (\text{D19})$$

and the conditional probability of having suffered some Pauli whilst knowing what the syndrome is would be

$$p(\mathbf{p}|\mathbf{s}) = \frac{\sum_{\xi \in \mathcal{L}_{\mathbf{p},\mathbf{s}}} P(\xi)}{\sum_{\xi' \in \mathcal{L}_{\mathbf{s}}} P(\xi')}. \quad (\text{D20})$$

Hence the effective logical noise channel after obtaining the syndrome $\mathbf{s}$ can be written as follows:

$$\mathcal{N}_{\text{logical},\mathbf{s}}(\cdot) = \sum_{\mathbf{p} \in \mathcal{P}} p(\mathbf{p}|\mathbf{s}) L(\mathbf{p})(\cdot) L(\mathbf{p})^\dagger, \quad (\text{D21})$$

where $L(\mathbf{p})$ is the logical Pauli operation corresponding to $\mathbf{p}$.

The coherent information (in terms of qubit per channel use) of the above logical noise channel (taking input state to be the maximally mixed state) after having obtained the syndrome of $\mathbf{s}$ assuming that the Pauli group has dimension d^2 is given by

$$\begin{aligned} R_{\mathbf{s}} &= \log_2(d) + \sum_{\mathbf{p}} p(\mathbf{p}|\mathbf{s}) \log_2(p(\mathbf{p}|\mathbf{s})) \\ &= \log_2(d) - H_2(p(\mathbf{p}|\mathbf{s})), \end{aligned} \quad (\text{D22})$$

which has a clear dependence on the syndrome. If we were to consider averaging this over the marginal distribution associated to achieving syndrome $\mathbf{s}$ we obtain

$$R = \log_2(d) - H_2(\mathbf{p}|\mathbf{s}), \quad (\text{D23})$$

where $H(\mathbf{p}|\mathbf{s})$ is the conditional entropy of $\mathbf{p}$ given $\mathbf{s}$. If one were to throw away the syndrome information, we are left

with the marginal distribution $p(\mathbf{p})$. The information gain which can be achieved using the syndrome information is the mutual information between $\mathbf{p}$ and $\mathbf{s}$ which is given by $I(\mathbf{p}; \mathbf{s}) = H_2(\mathbf{p}) - H_2(\mathbf{p}|\mathbf{s})$ which we know will always be non-negative.

To show that this is always an achievable rate, we note that $\mathcal{N}_{\text{logical}}$ can be understood as a classical mixture of Pauli channels written as

$$\mathcal{N}_{\text{logical}}(\cdot) = \int d\mathbf{s} \sum_{\mathbf{p}} p(\mathbf{p}, \mathbf{s}) L(\mathbf{p})(\cdot) L(\mathbf{p}) \otimes |\mathbf{s}\rangle\langle\mathbf{s}|, \quad (\text{D24})$$

for which the value R is its coherent information (taking input state to be the maximally mixed state) which is hence an achievable rate. We also know that our specific construction of polar codes will achieve this rate in the case of a single qudit Pauli noise.

1. Square GKP with Gaussian displacement noise

Let us consider the standard square GKP lattice defined by stabilizers

$$\hat{S}_1 = \exp(i\hat{q}\sqrt{2\pi d}), \quad \hat{S}_2 = \exp(-i\hat{p}\sqrt{2\pi d}), \quad (\text{D25})$$

and logical operators are $\hat{X}_L = S_2^{1/d}$ and $\hat{Z}_L = S_1^{1/d}$. We get stabilizer measurements $s_1 = \hat{q}\sqrt{d/2\pi} \bmod 1$ and $s_2 = -\hat{p}\sqrt{d/2\pi} \bmod 1$. Assuming a coherent displacement error to have occurred, the probability that closest point decoding results in a logical X^u is given by

$$\begin{aligned} p_1(u, s_1) &= p(s_1)p(u|s_1) \\ &= \frac{1}{\sigma\sqrt{d}} \sum_{l \in \mathbb{Z}} \exp\left(-\frac{\pi d}{\sigma^2} \left(\frac{s_1}{d} + \left(l + \frac{u}{d}\right)\right)^2\right) \\ &= \frac{1}{d} \theta_3\left(\pi \left(\frac{u + s_1}{d}\right), e^{-\pi\sigma^2/d}\right), \end{aligned} \quad (\text{D26})$$

where θ_3 is the Jacobi-theta function defined in Eq. (C3) and the marginal distribution for

$$p(s_1) = \theta_3(\pi s_1, e^{-d\pi\sigma^2}). \quad (\text{D27})$$

The joint distribution for $p_2(v, s_2) = p_1(v, -s_2)$ where Z^v is the phase error suffered due to the coherent displacement followed by closest point decoding. The symmetric relation between these two distributions along with the fact that u and v are not correlated comes from the use of a single mode square GKP code which is undergoing an isotropic displacement noise. We now have an achievable rate of

$$I_{d,\text{analog}}^{\text{sq}} = \log_2(d) + 2 \int ds_1 \sum_{u=0}^{d-1} p_1(u, s_1) \log_2\left(\frac{p_1(u, s_1)}{p(s_1)}\right). \quad (\text{D28})$$

Theorem D1. The rate $I_{d,\text{analog}}^{\text{sq}}$ is achievable by concatenation of a single mode square GKP code encoding a qudit of d levels (where d is prime) with a quantum polar code designed for the noise channel W_1 on both amplitude and phase.

Proof. We first note that the Pauli errors corresponding to k_1 and k_2 are independent of each other hence both can be treated to experience the noise channel described by W_1 . We begin with the same setup of freezing certain bits based on the sets $\mathcal{I}, \mathcal{A}, \mathcal{P}$ and $\mathcal{E}$. The first step we do is correct each individual GKP mode. This results in giving $\mathbf{k}_1$ and $\mathbf{k}_2$ as vectors of the stabilizer values and the state ending up in

$$|\Psi_1\rangle = \sum_{\mathbf{z}, \mathbf{u}, \mathbf{v}} p_{\mathbf{u}, \mathbf{v} | \mathbf{k}_1, \mathbf{k}_2}(\mathbf{z})_A X^{\mathbf{u}} Z^{\mathbf{v}} |G_N \mathbf{z}, \mathbf{k}_1, \mathbf{k}_2\rangle_B \otimes |\mathbf{u}, \mathbf{v}\rangle_E. \quad (\text{D29})$$

Now we can perform the decoding procedure the same way as in the proof of Theorem B2 whilst also using the outputs of $\mathbf{k}_1$ and $\mathbf{k}_2$ in the decoding procedure which essentially achieves the capacity for the classical channel W_1 as well as W_2 hence achieving $I_{d,\text{analog}} = 2I(W_1) - 1$ asymptotically [note that $I(W_1) = I(W_2)$ by the fact that $p_1(u, s_1) = p_2(u, -s_1)$]. ■

Theorem D2. For any given σ , there exists a d_0 such that for all prime $d > d_0$, the expression

$$\left| I_{d,\text{analog}}^{\text{sq}} - \log_2 \left(\frac{1}{\sigma^2 e} \right) \right| \quad (\text{D30})$$

can be made arbitrarily close to 0.

Since we know that d is prime and a large number we know that it is an odd number and so we can consider that the variable u can take on values from $-\lfloor d/2 \rfloor$ to $\lfloor d/2 \rfloor$ in steps of 1. For our convenience, we rewrite

$$\begin{aligned} I_{d,\text{analog}} &= \log_2 \left(\frac{1}{\sigma^2} \right) + 2 \int_{-1/2}^{1/2} ds_1 \\ &\times \left(\sum_u p_1(u, s_1) \log \left(\frac{p_1(u, s_1) \sigma \sqrt{d}}{p(s_1)} \right) \right). \end{aligned} \quad (\text{D31})$$

In the limit of $d \gg 1/\sigma^2$, we can note that

$$p(s_1) = 1 + \mathcal{O}(e^{-d\pi\sigma^2}) \quad (\text{D32})$$

since each syndrome value between $-1/2$ and $1/2$ becomes nearly equally likely. This follows from the fact that for a given syndrome, the set of displacements that

give the same syndrome are spaced apart by $\sqrt{2\pi/d}$ which gets increasingly fine as d increases. This nature can also be noted by the fact that $p(s_1)$ is a theta function.

In the summation of $p_1(u, s_1)$, the contribution from terms of $l = 0$ is significantly higher than any other term in the limit $d/\sigma^2 \gg 1$. Hence we can approximate the following truncated distribution of

$$p_{\text{lim}}(u, s_1) = \frac{1}{\sigma \sqrt{d}} \exp \left(-\frac{\pi d}{\sigma^2} \left(\frac{s_1}{d} + \frac{u}{d} \right)^2 \right). \quad (\text{D33})$$

To understand how far this might be from the actual probability distribution, we can integrate it to obtain

$$\int_{-1/2}^{1/2} ds_1 \sum_u p_{\text{lim}}(u, s_1) = \frac{1}{2} \left(\text{erf} \left(\frac{\sqrt{d\pi}}{2\sigma} \right) \right) \quad (\text{D34})$$

$$= 1 - e^{-\frac{d\pi}{4\sigma^2}} \left(\frac{2\sigma}{\pi \sqrt{d}} + \mathcal{O} \left((\sigma^2/d)^{-3/2} \right) \right), \quad (\text{D35})$$

which shows that the probability distribution is very close to $p_{\text{lim}}(u, s_1)$ when we have $e^{-\frac{d\pi}{4\sigma^2}} \rightarrow 0$. To accurately denote the differences in using this function instead of $p_1(u, s_1)$ we can write $p_1(u, s_1) = p_{\text{lim}}(u, s_1) + c(u, s_1)$ where $c(u, s_1)$ is appropriately defined for this. To upper bound $c(u, s_1)$ we note that it is comprised of the terms in Eq. (19) and if we individually maximize each term in the summation of l excluding $l = 0$ we can note that it corresponds to summing up $(2/\sigma \sqrt{d}) \exp(-\pi d(l - 1/2)^2/\sigma^2)$ from $l = 1$ to ∞ giving

$$|p_1(u, s_1) - p_{\text{lim}}(u, s_1)| = |c(u, s_1)| \leq \frac{\theta_2(0, e^{-d\pi/\sigma^2})}{\sigma \sqrt{d}}. \quad (\text{D36})$$

Note that $\theta_2(0, q) \leq 2q^{1/4}(1 - q)^{-1}$ when $q < 1$, and since we are working in the limit of small $e^{-\pi d/\sigma^2}$, we can hence note that $\sigma \sqrt{d} |c(u, s_1)| = \mathcal{O}(e^{-\frac{\pi d}{4\sigma^2}})$ as illustrated in Fig. 1(c).

For $d \gg \max\{1/\sigma^2, \sigma^2\}$, we expand $I_{d,\text{analog}}^{\text{sq}}$ as

$$\begin{aligned} I_{d,\text{analog}}^{\text{sq}} &= \log_2 \left(\frac{1}{\sigma^2} \right) + 2 \int_{-1/2}^{1/2} ds_1 \left(\left(\sum_u (p_{\text{lim}}(u, s_1) \right. \right. \\ &\quad \left. \left. + c(u, s_1)) \log \left(\sigma \sqrt{d} (p_{\text{lim}}(u, s_1) + c(u, s_1)) \right) \right) \right. \\ &\quad \left. - p(s_1) \log(p(s_1)) \right). \end{aligned} \quad (\text{D37})$$

We first note that $p(s_1) \log(p(s_1)) = \mathcal{O}(e^{-d\pi\sigma^2})$ since $p(s_1) = 1 + \mathcal{O}(e^{-d\pi\sigma^2})$. On further expansion we get

$$I_{d,\text{analog}}^{\text{sq}} = \log_2 \left(\frac{1}{\sigma^2} \right) + 2 \log_2(e)(w_1 + w_2 + w_3 + w_4) + \mathcal{O}(e^{-d\pi\sigma^2}), \quad (\text{D38})$$

where

$$w_1 = \sum_u \int_{-1/2}^{1/2} ds_1 p_{\text{lim}}(u, s_1) \ln(\sigma \sqrt{d} p_{\text{lim}}(u, s_1)), \quad (\text{D39})$$

$$w_2 = \sum_u \int_{-1/2}^{1/2} ds_1 c(u, s_1) \ln(\sigma \sqrt{d} p_{\text{lim}}(u, s_1)), \quad (\text{D40})$$

$$w_3 = \sum_u \int_{-1/2}^{1/2} ds_1 p_{\text{lim}}(u, s_1) \ln \left(1 + \frac{c(u, s_1)}{p_{\text{lim}}(u, s_1)} \right), \quad (\text{D41})$$

$$w_4 = \sum_u \int_{-1/2}^{1/2} c(u, s_1) \ln \left(1 + \frac{c(u, s_1)}{p_{\text{lim}}(u, s_1)} \right) ds_1, \quad (\text{D42})$$

and we will now proceed to show that w_1 gives a leading order term in the limit of large d while none of the other terms do.

Evaluating w_1 : We first evaluate the integral which is fairly easy since it is a Gaussian definite integral which simplifies to

$$\begin{aligned} w_1 &= \sum_u \frac{(e^{-\frac{\pi(1+2u)^2}{4d\sigma^2}}(2u+1) - e^{-\frac{\pi(2u-1)^2}{4d\sigma^2}}(2u-1))}{4\sigma\sqrt{d}} \\ &\quad + \sum_u \frac{1}{2} \left(\text{erf} \left(\frac{\sqrt{\pi}(1-2u/d)}{2\sigma\sqrt{d}} \right) \right. \\ &\quad \left. + \text{erf} \left(\frac{\sqrt{\pi}(1+2u/d)}{2\sigma\sqrt{d}} \right) \right) \\ &= \frac{1}{2} \text{erf} \left(\frac{\sqrt{d}\pi}{2\sigma} \right) + \frac{1}{2} \frac{e^{-\frac{\pi d}{4\sigma^2}} \sqrt{d}}{\sigma} \\ &= \frac{1}{2} + \exp \left(-\frac{\pi d}{4\sigma^2} \right) \frac{\sqrt{d}}{\sigma} \left(\frac{4-\pi}{2\pi} \right) \\ &= \frac{1}{2} + \mathcal{O}(d^{1/2} e^{-\frac{d\pi}{4\sigma^2}} \sigma^{-1}). \end{aligned} \quad (\text{D43})$$

Evaluating w_2 : We note that $|\log(\sigma \sqrt{d} p_{\text{lim}}(u, s_1))| = (\pi d/\sigma^2)(s_1/d - u/d)^2 \leq \pi d/\sigma^2$. Since $c(u, s_1)$ is positive,

this means $w_2 < 0$. We can bound it as

$$|w_2| \leq \sum_u \int_{-1/2}^{1/2} ds_1 c(u, s_1) \frac{\pi d}{\sigma^2} \quad (\text{D44})$$

$$= \frac{\pi d}{\sigma^2} \left(1 - \sum_u \int_{-1/2}^{1/2} ds_1 p_{\text{lim}}(u, s_1) \right) \quad (\text{D45})$$

$$= \frac{2\sqrt{d}}{\sigma} e^{-\frac{d\pi}{4\sigma^2}} \left(1 + \mathcal{O}((\sigma^2/d)^{-3/2}) \right), \quad (\text{D46})$$

hence giving $|w_2| = \mathcal{O}(d^{1/2} e^{-\frac{d\pi}{4\sigma^2}} \sigma^{-1})$.

Evaluating w_3 : We can first note that w_3 is positive in value. The minimum value of $\sigma \sqrt{d} p_{\text{lim}}(u, s_1)$ over the range of values that u and s_1 take is $e^{-\frac{d\pi}{4\sigma^2}}$ when $dx = \lfloor d/2 \rfloor$ (or $-\lfloor d/2 \rfloor$) and $s_1 = -1/2$ (or $1/2$). We can break the summation in w_3 into three intervals, $-\lfloor d/2 \rfloor \leq u < -\lfloor d/3 \rfloor$, $-\lfloor d/3 \rfloor \leq u \leq \lfloor d/3 \rfloor$ and $\lfloor d/3 \rfloor < u \leq \lfloor d/2 \rfloor$. When in the intervals $-\lfloor d/3 \rfloor \leq u \leq \lfloor d/3 \rfloor$ and $\lfloor d/3 \rfloor < u \leq \lfloor d/2 \rfloor$, $e^{-\frac{\pi d}{4\sigma^2}} \leq \sigma \sqrt{d} p_{\text{lim}}(u, s_1) \leq \alpha_1 e^{-\frac{\pi d}{9\sigma^2}}$ where $\alpha_1 \approx 1$. Hence in these intervals

$$\begin{aligned} &p_{\text{lim}}(u, s_1) \log \left(1 + \frac{c(u, s_1)}{p_{\text{lim}}(u, s_1)} \right) \\ &\leq \frac{1}{\sigma \sqrt{d}} \alpha_1 e^{-\frac{\pi d}{9\sigma^2}} \log \left(1 + \frac{\mathcal{O}(e^{-\frac{d\pi}{4\sigma^2}})}{e^{-\frac{d\pi}{4\sigma^2}}} \right) = \mathcal{O} \left(\frac{e^{-\frac{\pi d}{9\sigma^2}}}{\sigma \sqrt{d}} \right), \end{aligned} \quad (\text{D47})$$

and in the interval of $-\lfloor d/3 \rfloor \leq u \leq \lfloor d/3 \rfloor$, we have $\sigma \sqrt{d} p_{\text{lim}}(u, s_1) \geq \alpha_2 e^{-\frac{\pi d}{9\sigma^2}}$ where

$$\begin{aligned} &p_{\text{lim}}(u, s_1) \log \left(1 + \frac{c(u, s_1)}{p_{\text{lim}}(u, s_1)} \right) \\ &\leq p_{\text{lim}}(u, s_1) \log \left(1 + \frac{\mathcal{O}(e^{-\frac{d\pi}{4\sigma^2}})}{\alpha_2 e^{-\frac{\pi d}{9\sigma^2}}} \right) \\ &= p_{\text{lim}}(u, s_1) \log \left(1 + \mathcal{O}(e^{-\frac{5d\pi}{36\sigma^2}}) \right), \end{aligned} \quad (\text{D48})$$

and since we know that $e^{-\frac{d\pi}{\sigma^2}}$ is a small number this gives us $p_{\text{lim}}(u, s_1) \log(1 + c(u, s_1)/p_{\text{lim}}(u, s_1)) = \mathcal{O}(e^{-\frac{5d\pi}{36\sigma^2}}) p_{\text{lim}}(u, s_1)$. Summing over all these intervals, we

get

$$\begin{aligned}
 |w_3| &\leq \sum_{|u| \leq \lfloor \frac{d}{3} \rfloor} \mathcal{O}(e^{-\frac{5d\pi}{36\sigma^2}}) p_{\text{lim}}(u, s_1) \\
 &+ \sum_{\lfloor \frac{d}{3} \rfloor < |u| \leq \lfloor \frac{d}{2} \rfloor} \mathcal{O}\left(\frac{e^{-\frac{\pi d}{9\sigma^2}}}{\sigma\sqrt{d}}\right) \\
 &\leq \mathcal{O}(e^{-\frac{5d\pi}{36\sigma^2}}) + \mathcal{O}\left(d^{1/2} e^{-\frac{\pi d}{9\sigma^2}} \sigma^{-1}\right). \quad (\text{D49})
 \end{aligned}$$

Evaluating w_4 : as is already shown before, $\sigma\sqrt{d}p_{\text{lim}}(u, s_1) \geq e^{-\frac{d\pi}{4\sigma^2}}$. Using this we can place a bound on w_4 as

$$|w_4| \leq \alpha_3 \sum_u \int_{-1/2}^{1/2} ds_1 c(u, s_1) = \mathcal{O}(\sigma d^{-1/2} e^{-\frac{d\pi}{4\sigma^2}}). \quad (\text{D50})$$

Now we put together all of these together to get

$$\left| I_{d,\text{analog}}^{\text{sq}} - \log_2\left(\frac{1}{\sigma^2 e}\right) \right| \leq \mathcal{O}(d^{1/2} e^{-\frac{d\pi}{9\sigma^2}} \sigma^{-1}) + \mathcal{O}(e^{-d\pi\sigma^2}), \quad (\text{D51})$$

where this upper bound has a prefactor α_2 such that in the interval of $-[d/3] \leq u \leq [d/3]$, we have $\sigma\sqrt{d}p_{\text{lim}}(u, s_1) \geq \alpha_2 e^{-\frac{\pi d}{9\sigma^2}}$. The existence of fixed α_2 (with respect to d) holds true as long as we have $d > d_0$ such that $d_0/\sigma^2 \gg 1$. As such this upper bound will also tend to zero hence proving our claim. Therefore we have shown that

$$I_{d,\text{analog}}^{\text{sq}} = \log_2\left(\frac{1}{\sigma^2 e}\right) - \mathcal{O}(d^{1/2} e^{-\frac{d\pi}{9\sigma^2}} \sigma^{-1}) - \mathcal{O}(e^{-d\pi\sigma^2}). \quad (\text{D52})$$

2. Rectangular GKP with Gaussian displacement noise

We now examine the case of rectangular GKP with Gaussian displacement noise which would also result in the distributions associated to bit-flips and phase-flips being independent. Without loss of generality, let us assume the rectangle to be squeezed more along the $\hat{p}$ quadrature which would introduce a bias toward decreasing bit-flip errors and increasing phase-flip errors. The stabilizers are displacements defined as follows:

$$S_1 = \exp\left(i\hat{q}\frac{\sqrt{2\pi d}}{f}\right), \quad S_2 = \exp\left(-i\hat{p}f\sqrt{2\pi d}\right), \quad (\text{D53})$$

where logical operators $X_L = S_2^{1/d}$ and $Z_L = S_1^{1/d}$ and stabilizer measurements $s_1 = \hat{q}f\sqrt{d/2\pi} \bmod 1$ and

$s_2 = -\hat{p}(1/f)\sqrt{d/2\pi} \bmod 1$ where without loss of generality we assume $f > 1$. Assuming a coherent displacement error to have occurred, the probability that closest point decoding results in a logical X^u is given by

$$\begin{aligned}
 p_1(u, s_1) &= \frac{1}{f\sigma\sqrt{d}} \sum_{l \in \mathbb{Z}} \exp\left(-f^2 \frac{\left(s_1\sqrt{\frac{2\pi}{d}} + (dl+u)\sqrt{2\pi/d}\right)^2}{2f^2\sigma^2}\right) \\
 &= \frac{1}{d}\theta_3\left(\pi\left(\frac{u+s_1}{d}\right), e^{-\frac{\pi\sigma^2}{f^2d}}\right), \quad (\text{D54})
 \end{aligned}$$

and similarly the probability of a logical Z^v is given by

$$\begin{aligned}
 p_2(v, s_2) &= \frac{f}{\sigma\sqrt{d}} \sum_{l \in \mathbb{Z}} \exp\left(-\frac{1}{f^2} \frac{\left(s_2\sqrt{\frac{2\pi}{d}} - (dl+v)\sqrt{2\pi/d}\right)^2}{2\sigma^2}\right) \\
 &= \frac{1}{d}\theta_3\left(\pi\left(\frac{v-s_2}{d}\right), e^{-\frac{\pi f^2\sigma^2}{d}}\right). \quad (\text{D55})
 \end{aligned}$$

Note that this can be transformed exactly into the square case by only rescaling the value of σ . This gives the following achievable rate of

$$I_{d,\text{analog}}^{\text{rect},f}(\sigma) = \frac{1}{2}(I_{d,\text{analog}}^{\text{sq}}(\sigma f) + I_{d,\text{analog}}^{\text{sq}}(\sigma/f)). \quad (\text{D56})$$

We know that $I_{d,\text{analog}}^{\text{sq}} < \log_2(d)$ always and approaches $-\log_2(\sigma^2 e)$ as long as $d\sigma^2 \gg 1$. Hence if d is large enough such that also $d\sigma^2 \gg f^2$, then we have $I_{d,\text{analog}}^{\text{rect},f}(\sigma) \approx I_c(\sigma)$.

Assuming we are in a range of σ such that $I_{d,\text{analog}}^{\text{sq}}$ is concave in the range $[\sigma/f, \sigma f]$, we have

$$I_{d,\text{analog}}^{\text{rect},f}(\sigma) \leq I_{d,\text{analog}}^{\text{sq}}\left(\frac{\sigma}{2f}(f^2+1)\right) \leq I_{d,\text{analog}}^{\text{sq}}(\sigma), \quad (\text{D57})$$

where the first inequality follows from concavity of $I_{d,\text{analog}}^{\text{sq}}(\sigma)$ and the second follows from $(f^2+1) \geq 2f$ hence giving $\sigma \leq \sigma(f^2+1)/2f$. Hence biasing does not improve the rate for square GKP for certain ranges of sigma. As shown in Fig. 10, there is a range of values of σ for each d where $I_{d,\text{analog}}^{\text{sq}}$ is concave in σ which notably is far from the limit of where it approaches coherent information (which is notably convex in σ). Over this range, rectangular biasing (provably) provides no benefit.

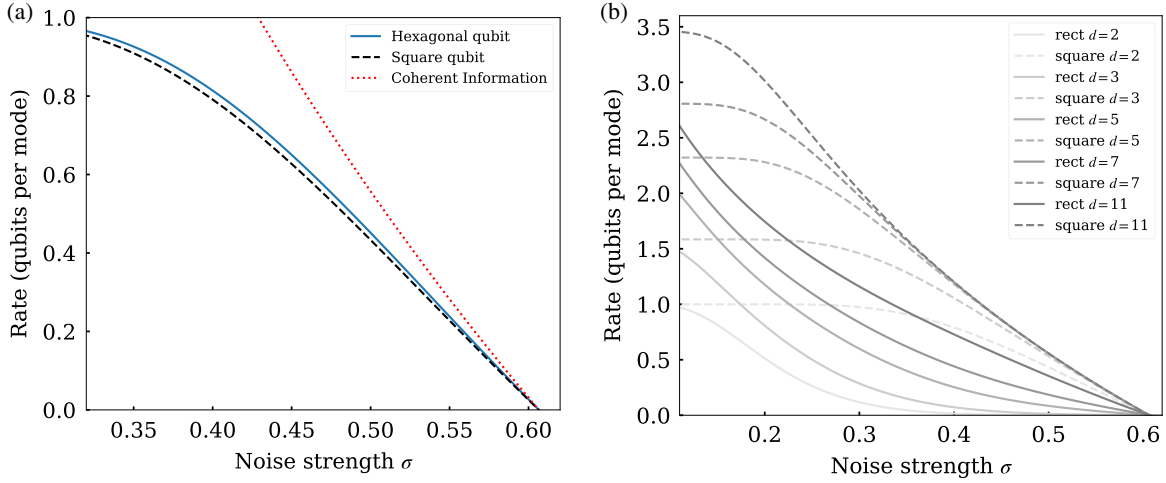


FIG. 10. (a) The achievable rates for a GKP hexagonal qubit slightly exceed that of a square GKP qubit as seen in this plot. (b) The achievable rates for rectangular GKP qubits with $f = 3$ (labeled as rect).

3. General single-mode GKP with Gaussian displacement noise

For any single-mode GKP, we would have the syndrome measurements given by s_1 and s_2 . In general, the probability of bit (or phase) flip would simultaneously depend on the values of s_1 and s_2 . Let us consider the event of logical error $X^u Z^v$ after correcting based on closest lattice point for syndrome values s_1 and s_2 . This would follow some probability distribution

$$p(u, v, s_1, s_2) \propto \sum_{\xi \in \mathcal{L}_{u,v,s_1,s_2}} \exp\left(-\frac{|\xi|^2}{2\sigma^2}\right) \quad (\text{D58})$$

where

$$\mathcal{L}_{u,v,s_1,s_2} = u\mathbf{x}_L + v\mathbf{z}_L + (MJ)^{-1}(\mathbf{s} - \mathbf{b}_0) + \mathcal{L} \quad (\text{D59})$$

where $\mathbf{x}_L$ ($\mathbf{z}_L$) is the displacement corresponding to a logical X (Z) operation. We also similarly would define $p(u, s_1, s_2) = \sum_v p(u, v, s_1, s_2)$. The polar codes associated to this would be obtained by looking at the two classical channels

$$\begin{aligned} W_1((z + u, s_1, s_2)|z) &= p(u, s_1, s_2), \\ W_2((x + v, u, s_1, s_2)|x) &= p(u, v, s_1, s_2) \end{aligned} \quad (\text{D60})$$

which gives an achievable rate of

$$\begin{aligned} R &= \log_2(d)(I(W_1) + I(W_2) - 1) \\ &= \log_2(d) \\ &\quad + \int ds_1 ds_2 \sum_{u,v} p(u, v, s_1, s_2) \log_2 \left(\frac{p(u, v, s_1, s_2)}{p(s_1, s_2)} \right). \end{aligned} \quad (\text{D61})$$

Note that the main difference here arises from the fact that we need to consider the closest point from $\mathbf{b}_0$ which would depend on both s_1 and s_2 which creates decision boundaries corresponding to the Voronoi cell.

APPENDIX E: FURTHER DETAILS ON NUMERICAL RESULTS

1. Symmetries between W_1 and W_2

For the memoryless noise channels W_1 and W_2 in Eq. (40), we have

$$W_1((z + u, s)|z) = W_2((x + u, -s)|x) \quad (\text{E1})$$

which can be directly observed by their definitions in terms of the functions $p_1(u, s_1)$ and $p_2(v, s_2)$. The errors of u and v are uncorrelated but follow symmetrically related distributions due to the isotropic nature of the underlying displacement noise. The polar code $(N, |\mathcal{A}|, \mathcal{A})_{d,\alpha}$ will have the exact same performance under noise model of W_1 and W_2 since they depict a bit-flip noise model of strength $p_1(u|s_1)$ and $p_2(v|s_2)$ respectively with probability $p(s_1)$ or $p(s_2)$ [these are the same distribution defined in Eq. (D27)]. Since $p(s_1) = p(-s_1)$, the chance of the noise being $p_1(u|s_1)$ is equally likely to that of $p(u|-s_1) = p_2(u|s_1)$, hence all relevant averaged parameters for W_1 and W_2 will be equal. Hence we will be only looking at W_1 for the rest of this discussion and understanding the phase-flip noise by looking at the kernel choice of $-\alpha$ instead of $+\alpha$ which will be the only change.

As discussed in Sec. IVC, to calculate the error probability of against noise channels W_1 or W_2 only requires the calculation of the error probability associated to communicating the all zero message u_i^N with $u_i = 0$. We now will highlight how these conditions allow us to claim that the if the classical polar codes $(N, |\mathcal{I} \cup \mathcal{P}|, \mathcal{I} \cup \mathcal{P})_{d,-\alpha}$

and $(N, |\mathcal{I} \cup \mathcal{A}|, \mathcal{I} \cup \mathcal{A})_{d,\alpha}$ are related by having each index $i \in (\mathcal{I} \cup \mathcal{P})$ correspond to $N - i \in (\mathcal{I} \cup \mathcal{A})$, they will have the same error probability.

Lemma E1. Consider a DMC $W: \mathbb{F}_d \rightarrow \mathcal{Y}$. If there exists a permutation $\pi: \mathcal{Y} \rightarrow \mathcal{Y}$ such that

$$W(y|x) = W(\pi(y)|-x) \quad (\text{E2})$$

and we consider the kernel that transforms $u_1, u_2 \in \mathbb{F}_d$ to $(u_1 + \alpha u_2, u_2)$ to transform the channels to

$$\begin{aligned} W^{-,\alpha}((y_1, y_2)|u_1) &= \frac{1}{d} \sum_{u_2 \in \mathbb{F}_d} W(y_1|u_1 + \alpha u_2) W(y_2|u_2) \\ W^{+,\alpha}((y_1, y_2, u_1)|u_2) &= \frac{1}{d} W(y_1|u_1 + \alpha u_2) W(y_2|u_2) \end{aligned} \quad (\text{E3})$$

then the following relation holds true for all $y_1, y_2 \in \mathcal{Y}$ and $u_1, u_2 \in \mathbb{F}_d$

$$\begin{aligned} W^{-,\alpha}((y_1, y_2)|u_1) &= W^{-,\alpha}((y_1, \pi(y_2))|u_1), \\ W^{+,\alpha}((y_1, y_2, u_1)|u_2) &= W^{+,\alpha}((\pi(y_1), y_2, -u_1)|u_2). \end{aligned} \quad (\text{E4})$$

Further for each of the new recombined channels, for all $\alpha \in \mathbb{F}_d \setminus \{0\}$ there is a similar symmetry as the input channel which is

$$\begin{aligned} W^{-,\alpha}((y_1, y_2)|u_1) &= W^{-,\alpha}((\pi(y_1), \pi(y_2))|-u_1) \\ W^{+,\alpha}((y_1, y_2, u_1)|u_2) &= W^{+,\alpha}((\pi(y_1), \pi(y_2), -u_1)|-u_2) \end{aligned} \quad (\text{E5})$$

Proof. We note that

$$\begin{aligned} \sum_{u_2} W(y_1|u_1 + \alpha u_2) W(y_2|u_2) \\ &= \sum_{u_2} W(y_1|u_1 - \alpha u_2) W(y_2|-u_2) \\ &= \sum_{u_2} W(y_1|u_1 - \alpha u_2) W(\pi(y_2)|u_2), \end{aligned} \quad (\text{E6})$$

and

$$\begin{aligned} W(y_1|u_1 + \alpha u_2) W(y_2|u_2) \\ &= W(\pi(y_1)|-u_1 - \alpha u_2) W(y_2|u_2) \end{aligned} \quad (\text{E7})$$

which when put together directly prove the above claim. $\blacksquare$

Lemma E2. For a channel $W: \mathbb{F}_d \rightarrow \mathcal{Y}$ let there exist another channel $W': \mathbb{F}_d \rightarrow \mathcal{Y}$ such that there exists a self-inverse permutation $\pi: \mathcal{Y} \rightarrow \mathcal{Y}$ satisfying for each x ,

$$W(y|x) = W'(\pi(y)|x), \quad (\text{E8})$$

which also satisfies

$$\begin{aligned} \forall x \left(\exists x' (W(y|x') > W(y|x)) \right) \\ \iff \exists x' (W(\pi(y)|x') > W(\pi(y)|x)). \end{aligned} \quad (\text{E9})$$

Defining $p_{\text{err}}(W)$ to be the error probability of one-shot maximum likelihood decoding and $Z_\Delta(W)$ to be the Bhattacharya distance [Eq. (A7)], we have

$$p_{\text{err}}(W) = p_{\text{err}}(W'), \quad Z_\Delta(W) = Z_\Delta(W'). \quad (\text{E10})$$

Proof. The probability of error is given by

$$p_{\text{err}}(W) = \frac{1}{d} \sum_{x \in \mathbb{F}_d} \sum_{y \in \mathcal{Y}} W(y|x) \mathbf{1}_{(y,x) \in \mathcal{E}_W} \quad (\text{E11})$$

where we define the error set $\mathcal{E}_W$ as

$$\mathcal{E}_W = \{(y, x) | y \in \mathcal{Y}, x \in \mathbb{F}_d, \exists x' \neq x, W(y|x') > W(y|x)\}. \quad (\text{E12})$$

Since we know that π is a permutation, from the definition of the above error set and the probability of error, it trivially follows that $p_{\text{err}}(W) = p_{\text{err}}(W')$. For the Bhattacharya distances, the permutation relation means that we can replace y with $\pi(y)$ in the summation

$$Z_\Delta(W) = \frac{1}{d} \sum_x \sum_{y \in \mathcal{Y}} \sqrt{W(y|x) W(y|x + \Delta)} \quad (\text{E13})$$

since it is over all of $\mathcal{Y}$. This proves both our claims. $\blacksquare$

We now make the observation that for a polar code using a SC decoder, the event of an error means that there for one of the indices $i \in \mathcal{A}$ suffered an error during decoding. Due to the order in which indices are decoded (which is in increasing order), we can see the event of failure for index i means that for all $j \in \mathcal{A}$ with $j < i$, the decoder succeeded but failed exactly at i . As such we can write down the error probability explicitly as

$$\begin{aligned} p_{\text{err}}^{(N, |\mathcal{A}|, \mathcal{A})_{d,\alpha}}(W) \\ &= \sum_{i \in \mathcal{A}} \left(\prod_{j \in \mathcal{A}, j < i} \left(1 - p_{\text{err}}(W_N^{(j), \alpha}) \right) \right) p_{\text{err}}(W_N^{(i), \alpha}). \end{aligned} \quad (\text{E14})$$

Claim E1. For the noise channels W_1 defined in Eq. (40), we consider their polarization under polar transform G_N with kernel choices of $+\alpha$ and $-\alpha$ respectively. We claim that

$$Z((W_1)_N^{(i),\alpha}) = Z((W_1)_N^{(i),-\alpha}) \quad (\text{E15})$$

$$p_{\text{err}}((W_1)_N^{(i),\alpha}) = p_{\text{err}}((W_2)_N^{(i),\alpha}) \quad (\text{E16})$$

and so as a consequence if the information sets are chosen such that $\forall i \in \mathcal{I} \cup \mathcal{P}, N-i \in \mathcal{I} \cup \mathcal{A}$ and the successive cancellation decoder decodes for W_2 in the reverse order for its input, these two polar codes have the exact same error probability.

Proof. Note that the permutation $\pi : \mathbb{F}_d \times [-1/2, 1/2) \rightarrow \mathbb{F}_d \times [-1/2, 1/2)$ defined by $\pi(e, s) = (-e, -s)$ gives us $W_1(\pi(y)|x) = W_1(y|x - x)$. This gives us the sufficient conditions for all the relations of Lemma E1 to hold. This condition also allows us to recursively claim that there exists an appropriate permutation $\pi^{(i,N)}$ composed of purely sign flips such that

$$\begin{aligned} (W_1)_N^{(i),\alpha}(\pi^{(i,n)}((y_1^N, u_i^{i-1}))|u_i) \\ = (W_1)_N^{(i),-\alpha}((y_1^N, u_i^{i-1})|u_i) \end{aligned} \quad (\text{E17})$$

which equivalently gives us the conditions for Lemma E2. The fact that this permutation is composed only of sign flips, the symmetry of the $p_1(u, s_1) = p(-u, -s_1)$ directly gives us the second condition in Lemma E1. The easiest way to see why is that it holds for $x = 0$ due to the sign flip nature and then can be extended to other values of x due to the symmetry of each $(W_1)_N^{(i)}$ with respect to its input. Putting all of this together along with Eq. (E14) gives us the condition that the $+\alpha$ and $-\alpha$ kernels will always have the exact same performance under this noise model. ■

Using this we can now trivially obtain that for our quantum polar code construction which reverses the order of dits in the X basis,

$$\begin{aligned} Z((W_1)_N^{(i)}) &= Z((W_2)_N^{(N-i)}), \\ p_{\text{err}}((W_1)_N^{(i)}) &= p_{\text{err}}((W_2)_N^{(N-i)}), \end{aligned} \quad (\text{E18})$$

and as such the error probabilities in Eq. (42) satisfy $P_{e,1} = P_{e,2}$.

2. Details on estimators for Bhattacharya distances

We have already highlighted that due to the permutation symmetry between the input dits, it is sufficient to do all sampling for the probability distribution of outputs generated by the all 0 input. As such for any channel W we define

the estimator

$$\hat{Z}_\Delta(W) = \frac{1}{M} \sum_{j=1}^M \sqrt{\frac{W(y_j|\Delta)}{W(y_j|0)}} \quad (\text{E19})$$

and we choose to sample y from the probability distribution $W(y|0)$. This is an unbiased estimator which satisfies

$$\begin{aligned} \mathbb{E}[\hat{Z}_\Delta(W)] &= Z_\Delta(W) \\ \text{Var}(\hat{Z}_\Delta(W)) &= \frac{(1 - Z_\Delta(W)^2)}{M}, \end{aligned} \quad (\text{E20})$$

where we denote Var to be the variance of this estimator. This means that if $Z_\Delta(W)$ is very close to 0, this is an estimator with a very high variance. Additionally, we can naively note that due to the tailed nature of the distribution for $Z_\Delta(W)$, it is significantly easier to underestimate its value if enough samples are not taken and also it is possible for its value to be highly skewed by a few larger observations. As such we avoid using these values to make any robust claims but do note that they can be useful in roughly identifying which of the indices i has $Z(W_N^{(i)}) \ll 1$ and which ones are close to 1.

To be able to reliably demonstrate the phenomenon of channel polarization we construct weak upper bounds for the actual values of $Z((W_1)_N^{(i)})$. For the sake of showing weak upper bounds, we can use the inequalities in Eq. (A10) to upper bound the values of $Z_\Delta(W_{2^m N}^{(i)})$ using the values of $Z_\Delta(W_N^{(i)})$. This would be naturally a significantly looser bound than explicitly calculating these values, but is useful to observe the phenomenon of polarization with fewer computational resources. However this upper bound is significantly weak in the sense that it may be possible that $Z(W_N^{(i)}) \ll 1$ but the value obtained by this upper bound $Z^{\text{upper}}(W_N^{(i)}) \rightarrow 1$. To see why, consider the case of $d = 2$ where the upper bound can be made finer as it satisfies $Z(W^-) \leq 2Z(W) - Z(W)^2$ [85]. The case of equality occurs for the binary erasure channel. Notably the binary erasure channel gives no information on likelihood of the input symbols in the event of the erasure. This method of finding Z^{upper} can be thought of as finding what the Bhattacharya distances behave like if the underlying channel was actually an erasure channel. In practice, the channels we work with are very far from erasure and the fact that the output is useful in giving different likelihoods for different symbols is key to why analog information provides us a provable benefit in the first place. We use Monte-Carlo methods for the actual estimation of $Z_\Delta(W_N^{(i)})$ which results in absolute errors scaling as $\mathcal{O}(1/\sqrt{M})$. We choose to take $M = 2^{20}$ and simulate for $N = 8$ and use this for finding these loose upper bounds for blocklengths of $N = 512$ to $N = 4096$ (as is plotted in Fig. 4).

3. Verification of error probabilities

Using Monte-Carlo sampling we estimate the values of $Z((W_1)_N^{(i)})$ which we consider as equal to $Z((W_2)_N^{(N-i)})$ which we use to construct our polar codes by trying to find the sets $\mathcal{I}, \mathcal{A}, \mathcal{P}$, and $\mathcal{E}$. These only provide a heuristic construction of the best possible quantum polar code due to stochasticity in the estimated values of the Bhattacharyya distances which cannot be made small without taking an unreasonable sample size. Hence, we cannot use this to actually provably upper bound the error probabilities. However, since we are working near a noise level where we only need to guarantee that it is upper bounded by a particular quantity, we can actually estimate the error probability and then verify that it is actually lower than the decreasing sequence which we choose to be $0.5 \times N^{-1/4}$. This choice is inspired from the original work on polar codes which proves rate achieving codes by upper bounding the error probability of the code by $N^{-1/4}$ for growing N [85]. This is a one-sided hypothesis testing task for the average of the Bernoulli random variable associated to a block-error. We find that using 2048 samples turns out to be more than enough to verify our claim that $P_{e,1} \leq 0.5 \times N^{-1/4}$ with much more than 99.9% confidence for each data point from $N = 2^9$ to $N = 2^{18}$ as can be seen in Fig. 11.

APPENDIX F: SELF-ORTHOGONAL CODES IN $GF(d^2)$

In this section we introduce stabilizer codes for qudits with a prime number of levels d , using self-orthogonal codes in $GF(d^2)$ using principles from Ref. [25]. We use the notation $[N, K]_d$ for a classical linear code which encodes K dits into N dits. Let us consider N qudit Pauli operators written as

$$X^{a^{(1)}} Z^{a^{(2)}} = \prod_{i=1}^N X^{a_i^{(1)}} Z^{a_i^{(2)}}, \quad (F1)$$

where $a_i^{(1)}, a_i^{(2)} \in \mathbb{F}_d$. The condition for two Pauli operators to commute is given by

$$[X^{a^{(1)}} Z^{a^{(2)}}, X^{b^{(1)}} Z^{b^{(2)}}] = 0 \iff a^{(1)} \cdot b^{(2)} = a^{(2)} \cdot b^{(1)} \pmod{d}, \quad (F2)$$

which can be understood as having the symplectic inner product between the two $2N$ -dimensional vectors $\mathbf{a} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b^{(1)} \\ b^{(2)} \end{pmatrix}$ being zero. If we consider $[[N, K]]_d$ stabilizer codes that encode K qudits in N qudits, we have $N - K$ stabilizers which we will consider to be from the N qudit Pauli group.

We consider some $[[N, K]]_d$ code defined by the stabilizer group $\mathcal{S} \subseteq \mathcal{P}_d^N$ where $\mathcal{P}_d^N$ is the N qudit Pauli

group. Let us consider the set $C \subseteq \mathbb{F}_d^{2N}$ constructed from the stabilizer group $\mathcal{S}$ defined by

$$C = \left\{ \begin{pmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \end{pmatrix} \middle| \hat{X}^{\mathbf{a}^{(1)}} \hat{Z}^{\mathbf{a}^{(2)}} \in \mathcal{S} \right\}. \quad (F3)$$

The set C is linearly additive since if $\mathbf{a}, \mathbf{b} \in C$ then so is $\alpha \mathbf{a} + \beta \mathbf{b} \in C$ for any $\alpha, \beta \in \mathbb{F}_d$. This means that C can be considered as a linear code which is $[2N, N - K]_d$ since it consists of $N - K$ linearly independent generators. We can define an inner product over $\mathbb{F}_d^{2N}$ by $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^{(1)} \cdot \mathbf{b}^{(2)} - \mathbf{a}^{(2)} \cdot \mathbf{b}^{(1)}$ which is a symplectic inner product. The relevance of this is that the two Pauli operators represented by $\mathbf{a}, \mathbf{b} \in \mathbb{F}_d^{2N}$ would commute iff $\langle \mathbf{a}, \mathbf{b} \rangle = 0$. Based on this we can define a symplectic dual code of C as

$$C^\perp = \{\mathbf{a} | \forall \mathbf{b} \in C, \langle \mathbf{a}, \mathbf{b} \rangle = 0\}, \quad (F4)$$

which would contain representations for all the Pauli operators which would give a logical error for the stabilizer code. For this to be a valid stabilizer code, we require that $C \subseteq C^\perp$ which is the condition of self-orthogonality. This ensures that all the stabilizers commute with each other.

We can now see that elements of C can also be considered to lie in the field of $\mathbb{F}_{d^2}^N$ since the Galois-field $GF(d^2)$ can be defined as a field extension of $GF(d)$ if d is a prime power. We will be restricting our discussion to having d be prime which lets us easily define $\mathbb{F}_d$ to be integers modulo d . However since Galois-fields require to have multiplication and division be defined along with addition and subtraction, this does not work for anything but d being prime. Hence one needs to extend the field of $GF(d)$ by expressing a polynomial using coefficients from $GF(d)$ which does not actually have a solution in $GF(d)$. The solutions of this polynomial will then be used as a basis to express elements in extensions of $GF(d)$. As an example if d is prime and $d \equiv 3 \pmod{4}$, there cannot exist any element which satisfies $x^2 = -1 \pmod{d}$ for any $x \in \mathbb{F}_d$. Hence we can express elements in $GF(d^2)$ as $a = a^{(1)} + \gamma a^{(2)}$ where $a^{(1)}, a^{(2)} \in \mathbb{F}_d$ and $\gamma \in \mathbb{F}_{d^2} \setminus \mathbb{F}_d$ satisfying $\gamma^2 = -1 \pmod{d}$.

The connection of classical self-orthogonal codes to stabilizer codes has been explored in various works, with the pivotal work exploring qubit stabilizer codes obtained from $GF(4)$ [86] and further works extending this for qudits using $GF(d^2)$ [25, 65, 66]. In the work by Ashikhmin and Knill [25], they show that for $\mathbf{a}, \mathbf{b} \in \mathbb{F}_{d^2}^N$ with $\mathbf{a} = \mathbf{a}^{(1)} + \gamma \mathbf{a}^{(2)}$ and $\mathbf{b} = \mathbf{b}^{(1)} + \gamma \mathbf{b}^{(2)}$ (here $\mathbf{a}^{(j)}, \mathbf{b}^{(j)} \in \mathbb{F}_d^N$ for $j = 1, 2$) the following holds true:

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b}^d &= \sum_i a_i b_i^d = 0 \pmod{d} \\ \implies \sum_i (a_i^{(1)} b_i^{(2)} - a_i^{(2)} b_i^{(1)}) &= 0 \pmod{d}. \end{aligned} \quad (F5)$$

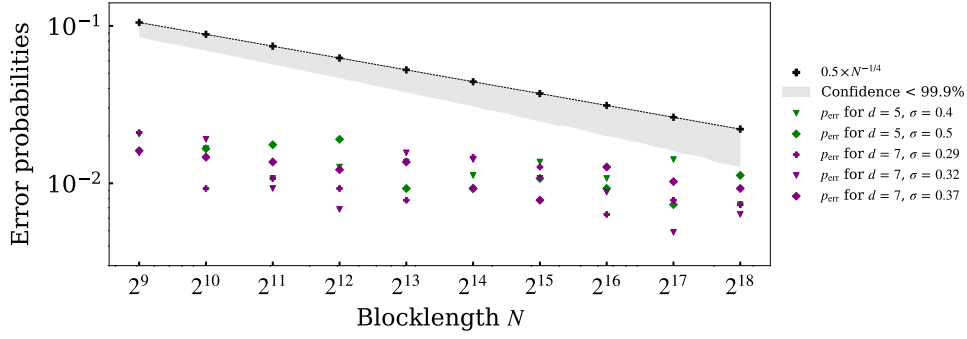


FIG. 11. Numerical verification for the trend of $p_{\text{err}} \leq 0.5 \times N^{-1/4}$. Here we have performed an estimation task for each p_{err} of the constructed polar codes. Our claim is that for all N in the sequence of codes, $p_{\text{err}} \leq 0.5 \times N^{-1/4}$ and we see that our estimated values satisfy this condition with larger than 99.9% confidence for each data point. The confidence here is defined for the one sided hypothesis task for whether $p_{\text{err}} \leq \delta$ with type-1 error $\leq 0.1\%$. While our explicit constructions do not demonstrate a strict downward trend, we are guaranteed by the martingale behavior of the Bhattacharya distances that we will always be able to continue along this sequence while still satisfying the condition of $p_{\text{err}} \leq 0.5 \times N^{-1/4}$. This sufficiently ensures $p_{\text{err}} \rightarrow 0$ as the rate approaches the value of $K/N \rightarrow I(W_1)$.

Hence they define the inner product $\mathbf{a}\mathbf{b} = \mathbf{a} \cdot \mathbf{b}^d$ which is also referred to as the Hermitian inner product in Ref. [66]. Using this a dual of a code in $\mathbb{F}_{d^2}^N$ can be defined. This inner product being zero provides a sufficient condition for two Pauli operators to commute. Note that this does not offer a necessary condition for commuting Pauli operators which can be trivially noted by seeing that in general for $\mathbf{a} \in \mathbb{F}_{d^2}^N$, $\mathbf{a}\mathbf{a}$ need not be zero. We note that in the case of $d = 3 \bmod 4$, the self-inner product takes on the nice form of

$$\mathbf{a}\mathbf{a} = \sum_i ((a^{(1)})^2 + (a^{(2)})^2) \bmod d, \quad (\text{F6})$$

which we will be making use of for proving our result in relation to achieving the capacity of pure-loss.

If we have a code $D \subseteq \mathbb{F}_{d^2}^N$ which is $[N, (N-K)/2]_d$ and the dual of it defined by

$$D^\perp = \{\mathbf{a} \in \mathbb{F}_{d^2}^N | \forall \mathbf{b} \in D, \mathbf{a}\mathbf{b} = 0\}, \quad (\text{F7})$$

which satisfies the self-orthogonality condition $D \subseteq D^\perp$, this implies the existence of a quantum code which is $[[N, K]]_d$. Every element in D satisfies self-orthogonality $\mathbf{a}\mathbf{a} = 0$. The number of non-zero elements in $\mathbb{F}_{d^2}^N$ which satisfy this property is given by

$$N_{\text{self}} = \frac{1}{d}(d^{2N} + (d-1)(-d)^N) - 1, \quad (\text{F8})$$

as shown in Ref. [68]. We now restate Lemma 7 from Ref. [25] which will be crucial in our analysis.

Lemma F1. Consider the set $\mathcal{T}$ consisting of all possible $[N, (N-K)/2]_d$ self-orthogonal codes. The number of codes in $\mathcal{T}$ which contains a given non-zero self-orthogonal vector $\mathbf{a}$ is independent of $\mathbf{a}$. Hence all non-zero

self-orthogonal vectors appear the same number of codes in $\mathcal{T}$.

Corollary F1. For any function $f : \mathbb{F}_{d^2}^N \rightarrow \mathbb{R}$

$$\frac{1}{|\mathcal{T}|} \sum_{D \in \mathcal{T}} \sum_{\mathbf{a} \in D} f(\mathbf{a}) = f(0) + \frac{d^{N-K} - 1}{N_{\text{self}}} \sum_{\mathbf{a}, \mathbf{a}\mathbf{a}=0, \mathbf{a} \neq 0} f(\mathbf{a}). \quad (\text{F9})$$

This follows from the basic-averaging lemma [87]. Note importantly that the sum is restricted to looking exactly at self-orthogonal $\mathbf{a}$. If we assume that d is a prime number such that $d = 3 \bmod 4$, we can assume a basis using $\gamma^2 + 1 = 0$ and so $\gamma = i$ and $\gamma_0 = 0$. This means that the self-orthogonal $\mathbf{a} = \mathbf{a}^{(1)} + i\mathbf{a}^{(2)}$ will satisfy $|\mathbf{a}^{(1)}|^2 + |\mathbf{a}^{(2)}|^2 = 0 \bmod d$.

APPENDIX G: ACHIEVING THE CAPACITY FOR PURE-LOSS

Lemma G1. The infidelity of an infinite-energy GKP code experiencing pure-loss (transmittance η) followed by the transpose recovery with an underlying symplectically integral lattice $\mathcal{L}$ encoding a finite dimension, can be upper bounded as

$$\epsilon \leq \frac{1}{4} \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus 0} e^{-\frac{\eta}{1-\eta} |\mathbf{x}|^2}. \quad (\text{G1})$$

The above lemma is taken from Ref. [17]. We now proceed to prove the main result.

Theorem G1. There exists a sequence of qudit stabilizer codes $[[N, K]]_d$ for d being prime with $d = 3 \bmod 4$, with

increasing N, d such that

$$\log_2(d) \frac{K}{N} = \log_2\left(\frac{\eta}{1-\eta}\right) - \tilde{\epsilon} \quad (\text{G2})$$

and $\tilde{\epsilon}$ can be made arbitrarily small simultaneously while the infidelity of this sequence of codes using transpose recovery after pure-loss of transmittance η converges to zero as $N, d \rightarrow \infty$ with $d \ln(d) \ll N \ll e^{\pi d \frac{\eta}{1-\eta}}$.

Proof. We first begin by noting a property of the summation ($g > 0$)

$$S_g = \sum_{z \in \mathbb{Z}^{2N} : |z|^2 \equiv 0 \pmod{d}} e^{-\pi g |z|^2/d}, \quad (\text{G3})$$

which is equivalent to

$$S_g = \frac{1}{d} \theta_3(e^{-\pi g/d})^{2N} + \frac{2}{d} \sum_{j=1}^{\lfloor (d-1)/2 \rfloor} \text{Re}(\theta_3(e^{-\pi g/d} \omega^j)^{2N}), \quad (\text{G4})$$

where $\omega = \exp(2\pi i/d)$. The use of ω ensures that any terms that have $|z|^2 \not\equiv 0 \pmod{d}$ will cancel out. Now using Eq. (C5) we note that

$$\theta_3(e^{-\pi g/d} \omega^j) = \theta_3(e^{-\pi \frac{d}{g-2ij}}) \sqrt{\frac{d}{g-2ij}}. \quad (\text{G5})$$

For any lattice, it is trivial that $|\Theta_{\mathcal{L}}(re^{i\theta})| \leq \Theta_{\mathcal{L}}(r)$ for $r \in \mathbb{R}^+$, $\theta \in [-\pi, \pi]$. Hence we get that

$$|\theta_3(e^{-\pi g/d} \omega^j)| \leq \sqrt{\frac{d}{g}} \left(\left(\frac{g^2}{g^2 + 4j^2} \right)^{1/4} \theta_3(e^{-\pi \frac{dg}{g^2 + 4j^2}}) \right). \quad (\text{G6})$$

We will now note that in the summation in Eq. G4, the central term of $j = 0$ dominates by an exponential amount compared to any $j \neq 0$. To check this, we first consider the following function:

$$h(t) = t^{1/4} \theta_3(e^{-\pi t}) = t^{-1/4} \theta_3(e^{-\pi/t}) = h(1/t). \quad (\text{G7})$$

It clearly follows that $h(t)$ has a local minima at $t = 1$ since its derivative equals zero, and perturbing around $t = 1$ only increases this function. To see why it is increasing for

$t > 1$, note that

$$\begin{aligned} h'(t) &= \frac{t^{-3/4}}{4} - \pi t^{1/4} e^{-\pi t} \theta_3'(e^{-\pi t}) \\ &= \frac{t^{-3/4}}{4} (1 - 4\pi t e^{-\pi t} \theta_3'(e^{-\pi t})), \end{aligned} \quad (\text{G8})$$

which we know is zero at $t = 1$. Also

$$\theta_3'(e^{-\pi t}) = 2 \sum_{n=1}^{\infty} n^2 e^{-(n^2-1)\pi t}, \quad (\text{G9})$$

which shows that $t > 1 \implies \theta_3'(e^{-\pi t}) < \theta_3'(e^{-\pi}) = e^{\pi}/4\pi$ which gives $h'(t) \geq t^{-3/4}(1 - te^{1-\pi t})$. Trivially, $te^{1-\pi t} < 1$ for all $t > 1$. Hence it follows that for $t > 1$, $h'(t) > 0$.

We substitute

$$t = \max \left\{ \frac{g^2 + 4j^2}{dg}, \frac{dg}{g^2 + 4j^2} \right\}, \quad (\text{G10})$$

which ensures $t > 1$ and since $h(t) = h(1/t)$ this gives

$$\frac{|\theta_3(e^{-\pi g/d} \omega^j)|}{\theta_3(e^{-\pi g/d})} \leq \left(\frac{t^{1/4} \theta_3(e^{-\pi t})}{\left(\frac{d}{g}\right)^{1/4} \theta_3(e^{-\pi \frac{d}{g}})} \right). \quad (\text{G11})$$

Let us suppose j always lies in $1 \leq 4j^2 \leq d^2 - d - g^2$. This already ensures that $t < d/g$ which means the above expression is strictly less than 1. In this range, the maximum value of $t^{1/4} \theta_3(e^{-\pi t})$ occurs at $4j^2$ being closest to $d^2 - d - g^2$ which gives

$$\begin{aligned} \frac{|\theta_3(e^{-\pi g/d} \omega^j)|}{\theta_3(e^{-\pi g/d})} &\leq \left(\left(1 - \frac{1}{d}\right)^{1/4} \frac{\theta_3(e^{-\pi \frac{d}{g}(1-d^{-1})})}{\theta_3(e^{-\pi \frac{d}{g}})} \right) \\ &\leq \left(1 - \frac{0.9}{d}\right)^{1/4}, \end{aligned} \quad (\text{G12})$$

when $e^{\pi d/g} \gg d$ since $\frac{\theta_3(e^{-\pi \frac{d}{g}(1-d^{-1})})}{\theta_3(e^{-\pi \frac{d}{g}})} \leq 1 + \mathcal{O}(e^{-\pi d/g})$.

This range already will contain all possible j since it only goes up till $(d-1)/2$. We can restrict to this choice of j without loss of generality since $\omega^j = \omega^{j \pmod{d}}$ where the modulo restricts it to the range $-(d-1)/2$ to $(d-1)/2$. We now fix $g = (1-\eta)/\eta$ and using Eqs. (C5) and (G12), we obtain the following:

$$\begin{aligned} S_{\frac{1-\eta}{\eta}} &\leq \frac{1}{d} d^N \left(\frac{\eta}{1-\eta} \right)^N \theta_3(e^{-\pi d(\frac{1-\eta}{\eta})})^{2N} \\ &\quad \times \left(1 + d \left(1 - \frac{0.9}{d} \right)^{\frac{N}{2}} \right). \end{aligned} \quad (\text{G13})$$

Now assuming we use a transpose recovery on a $[[N, K]]_d$ concatenated square GKP code, the infidelity can

be bounded by

$$4\epsilon \leq \sum_{\mathbf{x} \in \mathcal{L}^\perp \setminus \{0\}} e^{-\pi \frac{\eta}{1-\eta} |\mathbf{x}|^2}, \quad (\text{G14})$$

which using Eq. (C5) can be equivalently written as

$$4\epsilon \leq d^K \left(\frac{1-\eta}{\eta} \right)^N \Theta_{\mathcal{L}}(e^{-\pi \frac{1-\eta}{\eta}}) - 1, \quad (\text{G15})$$

since the $\det(\mathcal{L}) = d^K$ which is the total logical dimension.

Let us consider the set $\mathcal{T}$ composed of self-orthogonal codes $[N, (N-K)/2]_{d^2}$ which can be equivalently mapped to a set of stabilizer lattices $\mathcal{T}_{\mathcal{L}}$ composed of taking self-orthogonal codes D in $\mathcal{T}$ and mapping those to $2N$ -dimensional lattices. The explicit mapping is defined by

$$\mathcal{L}_D = \left\{ \frac{\mathbf{z}}{\sqrt{d}} \middle| \mathbf{z} = (\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) \in \mathbb{Z}^{2N}, \right. \\ \left. \text{with } ((\mathbf{z}^{(1)} + i\mathbf{z}^{(2)}) \bmod d) \in D \right\}, \quad (\text{G16})$$

which will represent a valid stabilizer lattice due to the self-orthogonality condition of $D \subseteq D^\perp$. Hence $\mathcal{T}_{\mathcal{L}}$ is composed of all the lattices $\mathcal{L}_D$ obtained from $D \in \mathcal{T}$. Due to the balanced nature of $\mathcal{T}$ over the set of self-orthogonal elements in $F_{d^2}^N$ (see Corollary F1), we get

$$\frac{1}{|\mathcal{T}_{\mathcal{L}}|} \sum_{\mathcal{L} \in \mathcal{T}_{\mathcal{L}}} \Theta_{\mathcal{L}}(e^{-\pi \frac{1-\eta}{\eta}}) \\ < \sum_{\mathbf{z} \in \mathbb{Z}^{2N} : \mathbf{z} \bmod d = 0} e^{-\pi |\mathbf{z}|^2 \frac{1-\eta}{\eta}} \\ + \frac{d^{N-K} - 1}{N_{\text{self}}} \sum_{\mathbf{z} \in \mathbb{Z}^{2N} : |\mathbf{z}|^2 \bmod d = 0} e^{-\pi |\mathbf{z}|^2 \frac{1-\eta}{\eta}} \\ < \theta_3(e^{-\pi d \frac{1-\eta}{\eta}})^{2N} + \frac{d^{N-K} - 1}{N_{\text{self}}} S_{\frac{1-\eta}{\eta}}, \quad (\text{G17})$$

where the first term comes from counting over all lattice points in $\sqrt{d}\mathbb{Z}^{2N}$ that are equivalent to the 0 codeword and the second term comes from counting all the points that are equivalent to some self-orthogonal codeword since we are using the basis $\{1, i\}$ for F_{d^2} (a valid choice since $d \equiv 3 \pmod{4}$). By representing $\mathbf{z} \bmod d = (\mathbf{a}^{(1)}, \mathbf{a}^{(2)})$ where the self-orthogonality for $\mathbf{a} = \mathbf{a}^{(1)} + i\mathbf{a}^{(2)}$ is equivalent to having $|\mathbf{z}|^2 \equiv 0 \pmod{d}$. There is over counting of codewords equivalent to 0 in the right-hand side, hence giving the inequality. Since we are evaluating an average over the whole set $\mathcal{T}_{\mathcal{L}}$, this means that there must exist some lattice

$\mathcal{L}_D \subseteq \mathcal{T}_{\mathcal{L}}$ which satisfies

$$\Theta_{\mathcal{L}_D}(e^{-\pi \frac{1-\eta}{\eta}}) < \theta_3(e^{-\pi d \frac{1-\eta}{\eta}})^{2N} + \frac{d^{N-K} - 1}{N_{\text{self}}} S_{\frac{1-\eta}{\eta}}. \quad (\text{G18})$$

Note that for any N and large enough d ,

$$\frac{d^{N-K} - 1}{N_{\text{self}}} < d^{-N-K+1}(1 + 1.1d^{-N+1}), \quad (\text{G19})$$

which when combined with Eq. (G13) gives the infidelity for $\mathcal{L}_D$ to satisfy

$$4\epsilon \leq \left(d^{K/N} \theta_3(e^{-\pi d \frac{1-\eta}{\eta}})^2 \left(\frac{1-\eta}{\eta} \right)^N \right)^N \\ + \left(1 + \frac{1.1}{d^{N-1}} \right) \left(1 + d \left(1 - \frac{0.9}{d} \right)^{\frac{N}{2}} \right) \theta_3(e^{-\pi d \frac{\eta}{1-\eta}})^{2N} \\ - 1. \quad (\text{G20})$$

Note that if the lattice $\mathcal{L}_D$ is associated to a GKP code with the rate R in per qubit, then we have $2^R = d^{K/N}$. For a certain rate to be achievable, we require that there must be a sequence of codes with certain encoding and decoding such that the infidelity in the limit of $N \rightarrow \infty$ is equal to zero.

Let us assume that we have the following guarantees in relating N and d which are

$$N \gg d \ln(d), \quad (\text{G21})$$

$$N \ll e^{\pi d \frac{\eta}{1-\eta}}. \quad (\text{G22})$$

Assuming that d is then large enough, we can always find a large enough d such that

$$\theta_3(e^{-\pi d \frac{1-\eta}{\eta}})^{-2} \geq 1 - 4.1e^{-\pi d \frac{1-\eta}{\eta}}, \quad (\text{G23})$$

$$(1 + 1.1d^{-N+1}) \left(1 + d \left(1 - 0.9d^{-1} \right)^{\frac{N}{2}} \right) \theta_3(e^{-\pi d \frac{\eta}{1-\eta}})^{2N} \\ \leq 1 + 4.1Ne^{-\pi d \frac{\eta}{1-\eta}}, \quad (\text{G24})$$

following which we force

$$d^{K/N} \leq \left(\frac{\eta}{1-\eta} \right) (1 - 4.1e^{-\pi d \frac{1-\eta}{\eta}})(1 - N^{-1+\delta}), \quad (\text{G25})$$

where $\delta > 0$. The residual factor of $(1 - N^{-1+\delta})$ is to ensure that raising to the power of N still makes the upper

bound to the infidelity approach zero. For this example we can choose $\delta = 1/2$. This gives

$$\begin{aligned} 4\epsilon &\leq \left(d^{K/N} \theta_3 \left(e^{-\pi d \frac{1-\eta}{\eta}} \right)^2 \left(\frac{1-\eta}{\eta} \right) \right)^N \\ &\quad + 4.1 N e^{-\pi d \frac{\eta}{1-\eta}} \\ &\leq 1.1 e^{-N^\delta} + 4.1 N e^{-\pi d \frac{\eta}{1-\eta}}, \end{aligned} \quad (\text{G26})$$

which using the relations above can be made arbitrarily small by appropriate choice of N and d . Specifically, we can define $N = \lfloor e^{d \frac{\eta}{1-\eta}} \rfloor$ for any given d . Then we can keep increasing d and N whilst defining K such that it satisfies Eq. (G25) and the value of ϵ for this sequence of codes will be decreasing and tend to zero with the achievable rate being

$$\begin{aligned} R_d &= \log_2(d^{K/N}) \\ &= \log_2 \left(\frac{\eta}{1-\eta} \right) - \mathcal{O}(e^{-\pi d \frac{1-\eta}{\eta}}) - \mathcal{O}(N^{-1+\delta}), \end{aligned} \quad (\text{G27})$$

which then clearly shows that this sequence of codes achieves the rate $\log_2(\eta/(1-\eta))$ which is the capacity of the loss channel. ■

-
- [1] C. Simon, Towards a global quantum network, *Nat. Photonics* **11**, 678 (2017).
 - [2] D. Leung, J. Oppenheim, and A. Winter, Quantum network communication—the butterfly and beyond, *IEEE Trans. Inf. Theory* **56**, 3478 (2010).
 - [3] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, in *STACS 2007*, edited by W. Thomas and P. Weil (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 610–621.
 - [4] A. Gandotra, Z. Wang, A. A. Clerk, and L. Jiang, Quantum communication over bandwidth-and-time-limited channels, *arXiv:2502.08831* [quant-ph].
 - [5] H. J. Kimble, The quantum internet, *Nature* **453**, 1023 (2008).
 - [6] C.-H. Wang, F. Li, and L. Jiang, Quantum capacities of transducers, *Nat. Commun.* **13**, 6698 (2022).
 - [7] H. Barnum, M. A. Nielsen, and B. Schumacher, Information transmission through a noisy quantum channel, *Phys. Rev. A* **57**, 4153 (1998).
 - [8] S. Lloyd, Capacity of the noisy quantum channel, *Phys. Rev. A* **55**, 1613 (1997).
 - [9] B. Schumacher and M. A. Nielsen, Quantum data processing and error correction, *Phys. Rev. A* **54**, 2629 (1996).
 - [10] A. S. Holevo and R. F. Werner, Evaluating capacities of bosonic Gaussian channels, *arXiv:quant-ph/9912067* [quant-ph].
 - [11] J. Harrington and J. Preskill, Achievable rates for the Gaussian quantum channel, *Phys. Rev. A* **64**, 062301 (2001).
 - [12] P. Leviant, Q. Xu, L. Jiang, and S. Rosenblum, Quantum capacity and codes for the bosonic loss-dephasing channel, *Quantum* **6**, 821 (2022).
 - [13] K. Noh, V. V. Albert, and L. Jiang, Quantum capacity bounds of Gaussian thermal loss channels and achievable rates with Gottesman-Kitaev-Preskill codes, *IEEE Trans. Inf. Theory* **65**, 2563 (2019).
 - [14] D. Gottesman, A. Kitaev, and J. Preskill, Encoding a qubit in an oscillator, *Phys. Rev. A* **64**, 012310 (2001).
 - [15] B. Royer, S. Singh, and S. M. Girvin, Encoding qubits in multimode grid states, *PRX Quantum* **3**, 010335 (2022).
 - [16] J. Conrad, J. Eisert, and F. Arzani, Gottesman-Kitaev-Preskill codes: A lattice perspective, *Quantum* **6**, 648 (2022).
 - [17] G. Zheng, W. He, G. Lee, K. Noh, and L. Jiang, Performance and achievable rates of the Gottesman-Kitaev-Preskill code for pure-loss and amplification channels, *PRX Quantum* **6**, 030314 (2025).
 - [18] P. Buser and P. Sarnak, On the period matrix of a Riemann surface of large genus (with an Appendix by J.H. Conway and N.J.A. Sloane), *Invent. Math.* **117**, 27 (1994).
 - [19] J. W. Harrington, Analysis of quantum error-correcting codes: Symplectic lattice codes and toric codes, Ph.D. thesis, California Institute of Technology, 2004.
 - [20] H. Cohn, A. Kumar, S. Miller, D. Radchenko, and M. Viazovska, Universal optimality of the e₈ and leech lattices and interpolation formulas, *Ann. Math.* **196**, 983 (2022).
 - [21] J. M. Renes, F. Dupuis, and R. Renner, Efficient polar coding of quantum information, *Phys. Rev. Lett.* **109**, 050504 (2012).
 - [22] N. Raveendran, N. Rengaswamy, F. Rozpędek, A. Raina, L. Jiang, and B. Vasić, Finite rate QLDPC-GKP coding scheme that surpasses the CSS Hamming bound, *Quantum* **6**, 767 (2022).
 - [23] K. Fukui, A. Tomita, and A. Okamoto, Analog quantum error correction with encoding a qubit into an oscillator, *Phys. Rev. Lett.* **119**, 180507 (2017).
 - [24] J. Conrad, J. Eisert, and J.-P. Seifert, Good Gottesman-Kitaev-Preskill codes from the NTRU cryptosystem, *Quantum* **8**, 1398 (2024).
 - [25] A. Ashikhmin and E. Knill, Nonbinary quantum stabilizer codes, *IEEE Trans. Inf. Theory* **47**, 3065 (2001).
 - [26] R. Sarkar and T. J. Yoder, The qudit Pauli group: non-commuting pairs, non-commuting sets, and structure theorems, *Quantum* **8**, 1307 (2024).
 - [27] For d being a positive integer power of a prime number, a Galois-field $GF(d)$ is a finite field with exactly d elements which supports addition, subtraction, multiplication and division (except by 0). For a quick intro to Galois-fields, we refer the reader to Ref. [88].
 - [28] G. Pantaleoni, B. Q. Baragiola, and N. C. Menicucci, Zak transform as a framework for quantum computation with the Gottesman-Kitaev-Preskill code, *Phys. Rev. A* **107**, 062611 (2023).
 - [29] M. H. Shaw, A. C. Doherty, and A. L. Grimsmo, Stabilizer subsystem decompositions for single- and multimode Gottesman-Kitaev-Preskill codes, *PRX Quantum* **5**, 010331 (2024).

- [30] D. Gottesman, A. Kitaev, and J. Preskill, Encoding a qubit in an oscillator, *Phys. Rev. A* **64**, 012310 (2001).
- [31] B. W. Walshe, B. Q. Baragiola, R. N. Alexander, and N. C. Menicucci, Continuous-variable gate teleportation and bosonic-code error correction, *Phys. Rev. A* **102**, 062411 (2020).
- [32] J. Hastrup and U. L. Andersen, Analysis of loss correction with the Gottesman-Kitaev-Preskill code, *Phys. Rev. A* **108**, 052413 (2023).
- [33] K. Noh and C. Chamberland, Fault-tolerant bosonic quantum error correction with the surface-Gottesman-Kitaev-Preskill code, *Phys. Rev. A* **101**, 012316 (2020).
- [34] K. Noh, C. Chamberland, and F. G. S. L. Brandão, Low-overhead fault-tolerant quantum error correction with the surface-GKP code, *PRX Quantum* **3**, 010315 (2022).
- [35] B. M. Terhal and D. Weigand, Encoding a qubit into a cavity mode in circuit QED using phase estimation, *Phys. Rev. A* **93**, 012315 (2016).
- [36] F. Rozpedek, K. Noh, Q. Xu, S. Guha, and L. Jiang, Quantum repeaters based on concatenated bosonic and discrete-variable quantum codes, *npj Quantum Inf.* **7**, 102 (2021).
- [37] M. Jafarzadeh, J. Conrad, R. N. Alexander, and B. Q. Baragiola, Logical channels in approximate Gottesman-Kitaev-Preskill error correction, [arXiv:2504.13383](https://arxiv.org/abs/2504.13383) [quant-ph].
- [38] M. H. Shaw, A. C. Doherty, and A. L. Grimsmo, Logical gates and read-out of superconducting Gottesman-Kitaev-Preskill qubits, [arXiv:2403.02396](https://arxiv.org/abs/2403.02396) [quant-ph].
- [39] M. Tomamichel, M. Berta, and J. M. Renes, Quantum coding with finite resources, *Nat. Commun.* **7**, 11419 (2016).
- [40] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
- [41] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, *Phys. Rev. A* **60**, 1888 (1999).
- [42] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, *Phys. Rev. A* **54**, 3824 (1996).
- [43] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, *Phys. Rev. A* **54**, 1098 (1996).
- [44] J. Fern and K. B. Whaley, Lower bounds on the nonzero capacity of Pauli channels, *Phys. Rev. A* **78**, 062335 (2008).
- [45] M. Lin and K. Noh, Exploring the quantum capacity of a Gaussian random displacement channel using Gottesman-Kitaev-Preskill codes and maximum likelihood decoding, *Phys. Rev. A* **111**, 052445 (2025).
- [46] E. Şaşıoğlu, E. Telatar, and E. Arikan, Polarization for arbitrary discrete memoryless channels, in *2009 IEEE Information Theory Workshop* (Taormina, Italy, 2009), pp. 144–148.
- [47] E. Arikan, Source polarization, in *2010 IEEE International Symposium on Information Theory* (IEEE, Austin, TX, USA, 2010), pp. 899–903.
- [48] E. Arikan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, *IEEE Trans. Inf. Theory* **55**, 3051 (2009).
- [49] R. Mori and T. Tanaka, Performance of polar codes with the construction using density evolution, *IEEE Commun. Lett.* **13**, 519 (2009).
- [50] S. H. Hassani, R. Mori, T. Tanaka, and R. L. Urbanke, Rate-dependent analysis of the asymptotic behavior of channel polarization, *IEEE Trans. Inf. Theory* **59**, 2267 (2013).
- [51] M. Mondelli, S. H. Hassani, and R. Urbanke, Construction of polar codes with sublinear complexity, in *2017 IEEE International Symposium on Information Theory (ISIT)* (Aachen, Germany, 2017), pp. 1853–1857.
- [52] I. Tal and A. Vardy, How to construct polar codes, [arXiv:1105.6164](https://arxiv.org/abs/1105.6164) [cs.IT].
- [53] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, On the construction of polar codes, in *2011 IEEE International Symposium on Information Theory Proceedings* (St. Petersburg, Russia, 2011), pp. 11–15.
- [54] R. Mori and T. Tanaka, Channel polarization on q-ary discrete memoryless channels by arbitrary kernels, in *2010 IEEE International Symposium on Information Theory* (Austin, TX, USA, 2010), pp. 894–898.
- [55] M. M. Wilde and S. Guha, Polar codes for classical-quantum channels, *IEEE Trans. Inf. Theory* **59**, 1175 (2013).
- [56] F. Dupuis, A. Goswami, M. Mhalla, and V. Savin, Purely Quantum Polar Codes, in *2019 IEEE Information Theory Workshop (ITW)* (Visby, Sweden, 2019), pp. 1–5.
- [57] F. Dupuis, A. Goswami, M. Mhalla, and V. Savin, Polarization of quantum channels using Clifford-based channel combining, *IEEE Trans. Inf. Theory* **67**, 2857 (2021).
- [58] A. Goswami, M. Mhalla, and V. Savin, Quantum Polarization of Qudit Channels, in *2021 IEEE International Symposium on Information Theory (ISIT)* (Melbourne, Australia, 2021), pp. 1487–1492.
- [59] A. Goswami, Quantum polar codes, Ph.D. thesis, Université Grenoble Alpes [2020–....], 2021.
- [60] T. A. Brun and M.-H. Hsieh, Entanglement-Assisted Quantum Error-Correcting Codes, in *Quantum Error Correction*, edited by Daniel A. Lidar and Todd A. Brun (Cambridge University Press, Cambridge, 2013), pp. 181–200.
- [61] G. Bowen, Entanglement required in achieving entanglement-assisted channel capacities, *Phys. Rev. A* **66**, 052313 (2002).
- [62] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, [arXiv:quant-ph/0106052](https://arxiv.org/abs/quant-ph/0106052) [quant-ph].
- [63] K. Oh, D. Kim, and J. Ha, Increasing Minimum Distance of Polar Codes with Outer Parity-Check Codes, in *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (Jeju, Korea (South), 2018), pp. 219–221.
- [64] F. Caruso and V. Giovannetti, Degradability of bosonic Gaussian channels, *Phys. Rev. A* **74**, 062307 (2006).
- [65] E. Rains, Nonbinary quantum codes, *IEEE Trans. Inf. Theory* **45**, 1827 (1999).
- [66] A. Ketkar, A. Klappenecker, S. Kumar, and P. Sarvepalli, Nonbinary stabilizer codes over finite fields, *IEEE Trans. Inf. Theory* **52**, 4892 (2006).
- [67] H.-A. Loeliger, Averaging bounds for lattices and linear codes, *IEEE Trans. Inf. Theory* **43**, 1767 (1997).

- [68] K. Feng and Z. Ma, A finite Gilbert-Varshamov bound for pure stabilizer quantum codes, *IEEE Trans. Inf. Theory* **50**, 3323 (2004).
- [69] B. L. Brock, S. Singh, A. Eickbusch, V. V. Sivak, A. Z. Ding, L. Frunzio, S. M. Girvin, and M. H. Devoret, Quantum error correction of qudits beyond break-even, *Nature* **641**, 612 (2025).
- [70] F. Schmidt and P. van Loock, Quantum error correction with higher Gottesman-Kitaev-Preskill codes: Minimal measurements and linear optics, *Phys. Rev. A* **105**, 042427 (2022).
- [71] Q. Xu, J. P. Bonilla Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasić, M. D. Lukin, L. Jiang, and H. Zhou, Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays, *Nat. Phys.* **20**, 1084 (2024).
- [72] Q. Xu, H. Zhou, G. Zheng, D. Bluvstein, J. P. B. Ataides, M. D. Lukin, and L. Jiang, Fast and parallelizable logical computation with homological product codes, [arXiv:2407.18490](https://arxiv.org/abs/2407.18490) [quant-ph].
- [73] D. J. Williamson and T. J. Yoder, Low-overhead fault-tolerant quantum computation by gauging logical operators, [arXiv:2410.02213](https://arxiv.org/abs/2410.02213) [quant-ph].
- [74] E. Swaroop, T. Jochym-O'Connor, and T. J. Yoder, Universal adapters between quantum LDPC codes, [arXiv:2410.03628](https://arxiv.org/abs/2410.03628) [quant-ph].
- [75] A. Cowtan, Z. He, D. J. Williamson, and T. J. Yoder, Parallel logical measurements via quantum code surgery, [arXiv:2503.05003](https://arxiv.org/abs/2503.05003) [quant-ph].
- [76] Z. He, A. Cowtan, D. J. Williamson, and T. J. Yoder, Extractors: QLDPC architectures for efficient Pauli-based computation, [arXiv:2503.10390](https://arxiv.org/abs/2503.10390) [quant-ph].
- [77] S. K. Borah, A. K. Pradhan, N. Raveendran, M. Pacenti, and B. Vasic, Fault tolerant decoding of QLDPC-GKP codes with circuit level soft information, [arXiv:2505.06385](https://arxiv.org/abs/2505.06385) [quant-ph].
- [78] J.-L. Kim and J. Walker, Nonbinary quantum error-correcting codes from algebraic curves, *Discrete Math.* **308**, 3115 (2008).
- [79] G. G. La Guardia, Asymmetric quantum Reed-Solomon and generalized Reed-Solomon codes, *Quantum Inf. Process.* **11**, 591 (2012).
- [80] L. Golowich and V. Guruswami, Asymptotically good quantum codes with transversal non-Clifford gates, [arXiv:2408.09254](https://arxiv.org/abs/2408.09254) [quant-ph].
- [81] W. Ebeling, *Lattices and Codes*, Advanced Lectures in Mathematics (Springer-Verlag, New York, NY, USA, 1994).
- [82] M. Subramanian, Github repository achievable-rates-SQ-GKP, 2025, <https://github.com/mahadevans2432/achievable-rates-sq-gkp/>.
- [83] E. Arikan and E. Telatar, On the rate of channel polarization, in *2009 IEEE International Symposium on Information Theory* (Seoul, Korea (South), 2009), pp. 1493–1495.
- [84] N. Elkies, Math 272y: Rational lattices and their theta functions (Fall, 2019), <https://people.math.harvard.edu/~elkies/M272.19/index.html>, accessed 31-10-2025.
- [85] E. Arikan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, *IEEE Trans. Inf. Theory* **55**, 3051 (2009).
- [86] A. Calderbank, E. Rains, P. Shor, and N. Sloane, Quantum error correction via codes over GF(4), *IEEE Trans. Inf. Theory* **44**, 1369 (1998).
- [87] H.-A. Loeliger, Averaging bounds for lattices and linear codes, *IEEE Trans. Inf. Theory* **43**, 1767 (1997).
- [88] R. E. Blahut, *Algebraic Codes for Data Transmission* (Cambridge University Press, Cambridge, 2003).