

Probabilistic channel distillation via indefinite causal order

Spiros Kechrimparis^{1,*}, James Moran^{2,†} and Hyukjoon Kwon^{1,2,‡}

¹*School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, South Korea*

²*Quantum Universe Center, Korea Institute for Advanced Study, Seoul 02455, South Korea*



(Received 12 February 2025; accepted 4 November 2025; published 5 December 2025)

The quantum switch has been widely studied as a prototypical example of indefinite causal order in quantum information processing. However, the potential advantages of utilizing more general forms of indefinite causal orders remain largely unexplored. We study higher-order switches, which involve concatenated applications of the quantum switch, and we demonstrate that they provide a strict advantage over the conventional quantum switch in the task of quantum channel distillation. Specifically, we show that higher-order quantum switches enable the probabilistic distillation of any qubit Pauli channel into the identity channel with nonzero probability. This capability contrasts with the conventional quantum switch, which allows only a limited set of Pauli channels to be distilled with nonzero probability. We observe that, counterintuitively, the distillation rate generally increases the noisier the channel is. We fully characterize the asymptotic distillation rates of higher-order superswitches for qubit Pauli channels. Finally, we prove a no-go result for multiqubit generalizations.

DOI: [10.1103/mvrg-yj3l](https://doi.org/10.1103/mvrg-yj3l)

I. INTRODUCTION

Emerging from the context of quantum gravity [1,2], it was observed that quantum theory is consistent with operations that are not applied in a definite causal order [3,4], an observation that has sparked intense research since. The *quantum switch* [4], the prototypical example of a supermap [5,6] with indefinite causal order, has since found application in a variety of topics ranging from quantum computation [7,8], quantum metrology [9], to quantum work extraction [10]. Advantages for classical and quantum communication by using indefinite causal order have been shown in Refs. [11–21]. Since then, various experiments have been performed that demonstrate these findings; for a review, see Ref. [22]. In Ref. [23], it was shown that noiseless communication is possible deterministically through an entanglement-breaking channel that is unique, up to unitary equivalence, by inputting two copies of the channel in the quantum switch. Previously, probabilistic perfect transmission of quantum information was shown to be possible with the quantum switch for some Pauli channels [12].

Higher-order generalizations of the quantum switch have been introduced [15–21]. The *n-switch* superposes the orderings of *n* channels through coherent control of all possible *n!* orders [15,16]. Superposition of the *N* cyclic orders of *N*

depolarizing channels has also shown advantages in communication [18]. The switch of switches has also been considered in Refs. [20,21] and has demonstrated advantages in quantum communication and state discrimination. Although they have demonstrated advantages in some instances, in most cases these improvements are not always clear-cut. In addition, the coherence needed to implement higher-order switches scales badly, in general, further diminishing the returns and deeming their usefulness questionable. In this work, we first report on a task that certain types of higher-order switches can perform, while the quantum switch cannot, thus demonstrating that they can have a clear advantage over the quantum switch. The task at hand is exact probabilistic channel distillation, which by analogy to other resource distillation scenarios [24–37] consists of turning multiple copies of a noisy channel into the identity or a unitary channel, with some nonzero probability.

We first notice that the quantum switch cannot perform channel distillation except for channels from a subset of measure zero in the set of unital channels. We then show that by using higher-order switches, and specifically switches of switches, probabilistic distillation is possible with *any* unital channel in dimension 2. Counterintuitively, we find that the noisier a channel is, the better it performs in this distillation protocol.

The quantum switch and some *n*-switches for small *n* have been implemented [22,38], and thus our results can be employed to achieve noiseless classical and quantum communication in $d = 2$ with any channel. We remark that as such experiments require more uses of the channels in comparison to the theoretical indefinite causal order setting, they are often considered simulations instead of direct realizations. For the intricacies and the debate on experimental implementations, see Ref. [22]. From a foundational perspective, and in view of the Choi-Jamiołkowski isomorphism [39,40], our results have

*Contact author: skechrimparis@gmail.com

†Contact author: jamesmoran@kias.re.kr

‡Contact author: hjkwon@kias.re.kr

implications for entanglement sharing between two parties, where the first party prepares an entangled state, keeps one part of the system, and sends the second part to the other party over a quantum channel. This task is impossible over a zero quantum capacity channel. With access to higher-order switches, any unital channel in dimension $d = 2$ can be used to distribute entanglement with nonzero probability, including all channels from the set of entanglement-breaking channels, enabling applications that would otherwise be impossible.

II. QUANTUM SWITCH AND SUPERSWITCHES

The quantum switch [4] is a supermap that superposes the ordering between the actions of two channels $\mathcal{E}$ and $\mathcal{F}$, conditioned to a two-dimensional switch state ω . The resulting channel is defined as

$$S_\omega(\mathcal{E}, \mathcal{F}) = \sum_{i,j} S_{ij}(\rho \otimes \omega) S_{ij}^\dagger, \quad (1)$$

with $S_{ij} = E_i F_j \otimes |0\rangle\langle 0|_C + F_j E_i \otimes |1\rangle\langle 1|_C$. Here, E_i and F_j denote the Kraus operators of the channels $\mathcal{E}$ and $\mathcal{F}$, respectively, and $|0\rangle\langle 0|_C$ and $|1\rangle\langle 1|_C$ denote the projectors acting on the ancilla state. S_{ij} are Kraus operators satisfying $\sum_{i,j} S_{ij}^\dagger S_{ij} = 1$ when both $\mathcal{E}$ and $\mathcal{F}$ are Pauli channels, represented as $\mathcal{P}_{\vec{p}}(\rho) = p_0 \rho + p_1 X \rho X + p_2 Y \rho Y + p_3 Z \rho Z$ with $\vec{p} = (p_1, p_2, p_3)$ and $p_0 = 1 - p_1 - p_2 - p_3$ [4,23]. In particular, by taking the ancilla state to be $\omega = |+\rangle\langle +|$ with $| \pm \rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, we obtain

$$S_\omega(\mathcal{E}, \mathcal{F}) = C_+(\mathcal{E}, \mathcal{F})(\rho) \otimes |+\rangle\langle +| + C_-(\mathcal{E}, \mathcal{F})(\rho) \otimes |-\rangle\langle -|, \quad (2)$$

where

$$C_\pm(\mathcal{E}, \mathcal{F})(\rho) = \sum_{i,j} [E_i, F_j]_\pm \rho [E_i, F_j]_\pm^\dagger,$$

with $[A, B]_\pm = AB \pm BA$ denoting the anticommutator and commutator. $C_\pm(\mathcal{E}, \mathcal{F})(\rho)$ can be expressed as probabilities $q_\pm^{(1)}$, multiplied by channels $\mathcal{E}_\pm^{(1)}(\rho) = C_\pm(\mathcal{E}, \mathcal{F})(\rho)/q_\pm^{(1)}$, where $q_\pm^{(1)} = \text{tr} C_\pm(\mathcal{E}, \mathcal{F})(\rho)$. Consequently, a measurement of the ancilla state in the basis $\{|+\rangle, |-\rangle\}$ results in the separation of the two channels. The case with $\mathcal{E} = \mathcal{F}$ is most commonly studied, which we also assume in this work.

Higher-order generalizations of the quantum switch have also been defined [15–20]. A particular case is switches of switches, also known as superswitches. An instance with synchronized ordering on the inside switches was studied in Ref. [20], while the general case was introduced and studied in Ref. [21] and applied to the problem of quantum state discrimination. The n -order superswitch effectively evaluates all possible nested expressions of commutators and anticommutators of the Kraus operators of the channels. We denote superswitches by SWITCH_n , with n denoting their order. SWITCH_n is obtained by applying the switch operation in Eq. (2) to the two lower-order superswitches SWITCH_{n-1} (see Fig. 1). The case $n = 0$ corresponds to the application of no superswitch, i.e., just the channel, and $n = 1$ is the case of the quantum switch. To implement SWITCH_n , 2^n copies of the channel $\mathcal{E}$ and $2^n - 1$ ancilla qubits are needed. The role of

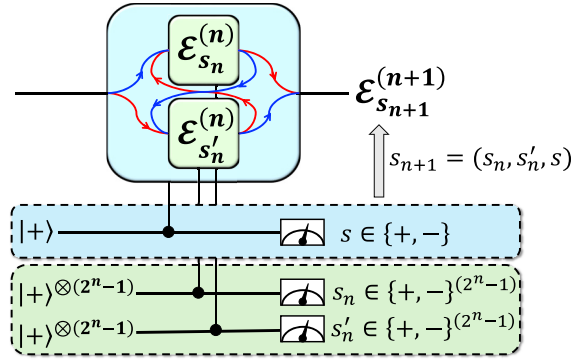


FIG. 1. The $(n + 1)$ -order superswitch is generated by superposing the ordering two n -order superswitches. The ancilla qubits (green) control the ordering of channels inside the n -order superswitches, while another ancilla (blue) controls their ordering.

the ancillas is to control the ordering of all nested switches. We remark that SWITCH_n can be seen as a restricted form of the n -switch (see Appendix A); however, owing to its iterative definition, it can be readily evaluated.

At each order, the action of SWITCH_n can be effectively described by the tuples $\Omega_{s_n}^{(n)} = \{q_{s_n}^{(n)}, \mathcal{E}_{s_n}^{(n)}\}$, with $\mathcal{E}_{s_n}^{(n)}$ denoting channels and $q_{s_n}^{(n)}$ their probability of occurrence; s_n denotes a string of length $2^n - 1$ consisting of “ $\pm$ ” outcomes corresponding to the measurements on the ancilla qubits. A superswitch of any order n can be evaluated from the previous order $n - 1$ (see Fig. 1). Explicitly, we find for the tuples $\Omega_{s_n}^{(n)} = \{q_{s_n}^{(n)}, \mathcal{E}_{s_n}^{(n)}\}$,

$$\begin{aligned} \mathcal{E}_{s_n}^{(n)} &= C_s(\mathcal{E}_{s_{n-1}}^{(n-1)}, \mathcal{E}_{s'_{n-1}}^{(n-1)})/q_{s_n}^{(n)}, \\ q_{s_n}^{(n)} &= \text{tr}[C_s(\mathcal{E}_{s_{n-1}}^{(n-1)}, \mathcal{E}_{s'_{n-1}}^{(n-1)})(\rho)], \end{aligned} \quad (3)$$

where $s_n = s_{n-1}s'_{n-1}s$ denotes the concatenation of strings s_{n-1} , s'_{n-1} , and s . At order $n = 0$, we only have the channel, and thus the initial conditions are $q^{(0)} = 1$ and $\mathcal{E}^{(0)} = \mathcal{E}$. Iterating once, we get the two channels of the conventional quantum switch, while a second iteration gives the first higher-order generalization $\Omega_{\mu\nu s}^{(2)} = \{q_{\mu\nu s}^{(2)}, \mathcal{E}_{\mu\nu s}^{(2)}\}$ consisting of eight channels and their respective probabilities of occurrence, and where μ, ν are strings of “ $\pm$ ” of length 1. In Appendix B, we list all the expressions $\Omega_{\mu\nu s}^{(2)}$ as well as the general expression for the n -order superswitch.

III. QUANTUM CHANNEL DISTILLATION VIA HIGHER-ORDER QUANTUM SWITCHES

Resource distillation is the task of converting k less-resourceful copies of a certain resource into l copies with higher resource content, where $k \geq l$. Examples include protocols to distil entanglement, coherence, and nonlocality [24–37]. Such a task can be exact or approximate depending on whether the target resource is achieved exactly or up to a certain approximation. In addition, such protocols can be deterministic or probabilistic, depending on whether the conversion is successful with certainty or only with some nonzero probability. In this work, we consider the task of exact distillation of multiple copies of a given channel $\mathcal{E}$

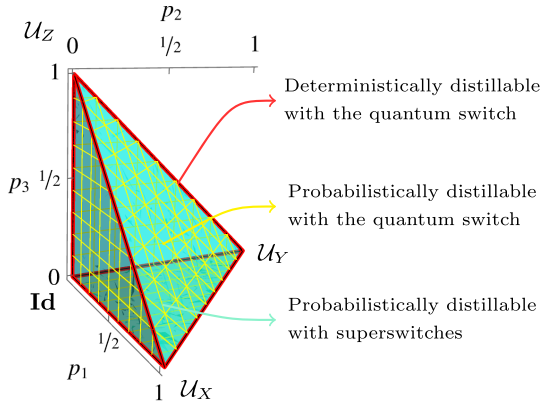


FIG. 2. The tetrahedron of the set of Pauli channels. On the edges are channels that can be distilled with the quantum switch deterministically (red lines), while channels on the faces can be distilled probabilistically (yellow meshed surfaces). In the interior are channels that can only be distilled with superswitches (cyan). $\mathcal{U}_P$ denotes a unitary channel $P \cdots P^\dagger$ with $P = X, Y, Z$.

into the identity or a unitary channel $\mathcal{U}$. In the latter case, applying the inverse unitary recovers the identity channel as well. Thus, we will consider exact probabilistic channel distillation protocols. Our proposed channel distillation protocols employ SWITCH_n , which require 2^n copies of the channel $\mathcal{E}$ in order to implement.

Definition 1. Let $D^{(n)}$ be the collection of strings of $2^n - 1$ plus or minus outcomes such that the corresponding channel is the identity or a unitary channel. That is, with $\tilde{s} \in D^{(n)}$ we either have $\mathcal{E}_{\tilde{s}}^{(n)} = \text{Id}$ or $\mathcal{E}_{\tilde{s}}^{(n)} = \mathcal{U}$, where $\mathcal{U}(\rho) = U\rho U^\dagger$ denotes an arbitrary unitary channel. Then, the *distillation rate* for a channel $\mathcal{E}$, $\mathcal{R}_{\mathcal{E}}^{(n)} \in [0, 1]$, of a protocol based on SWITCH_n is given by

$$\mathcal{R}_{\mathcal{E}}^{(n)} = \sum_{\tilde{s} \in D^{(n)}} q_{\tilde{s}}^{(n)}. \quad (4)$$

In Ref. [23], it was shown that the channel $\mathcal{E}_{XY} = \frac{1}{2}(X\rho X + Y\rho Y)$ is the unique channel that is both entanglement breaking and therefore useless for quantum communication, but leads to *perfect* communication with unit probability when combined with itself in the quantum switch. In our setting, this translates to the statement that there is a unique entanglement-breaking channel, up to unitary transformations, that can be exactly purified deterministically with the quantum switch. More generally, any channel with zero portion of the identity, $p_0 = 0$, can be probabilistically purified with the quantum switch.

By considering probabilistic protocols, more channels can be purified. As the set of Pauli channels $\mathcal{P}_{\vec{p}}(\rho) = p_0\rho + p_1X\rho X + p_2Y\rho Y + p_3Z\rho Z$ with $\vec{p} = (p_1, p_2, p_3)$ forms a tetrahedron in (p_1, p_2, p_3) space (see Appendix C), the set of channels that can be distilled by the quantum switch lies at the boundary of the tetrahedron, that is, the four faces and the channels at the edges. Inside the tetrahedron, no channel can be distilled with the quantum switch (see Fig. 2). We note that even though the three axes of the tetrahedron consist of channels that cannot be distilled by direct application of the quantum switch (or even the protocols we will

propose), a simple application of one of the Pauli matrices $\sigma_i \in \{\sigma_1, \sigma_2, \sigma_3\} = \{X, Y, Z\}$ with the same index i as one of the zero p_i 's in the definition of the channel, before inputting in the quantum switch, maps it to a channel with nonzero p_i 's. This straightforward preprocessing is a technique that has also been employed in Refs. [41,42].

IV. DISTILLATION PROTOCOL WITH SWITCH₂

Our first result shows that SWITCH_2 allows purification of all channels inside the tetrahedron.

Result 1. Any qubit Pauli channel inside the tetrahedron can be exactly distilled to the identity with a nonzero probability when four copies of the channel are combined in the second-order superswitch. Moreover, a Pauli channel $\mathcal{E} = \mathcal{P}_{\vec{p}}$ with $p_i \neq 0 \forall i$ has distillation probability:

$$\mathcal{R}_{\mathcal{E}}^{(2)} = q_{---+}^{(2)} = 4(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2). \quad (5)$$

The proof is given in Appendix D. We note that this probability is nonzero, i.e., $q_{---+}^{(2)} > 0$, for all channels except for the trivial case of $\mathcal{E}_{\vec{p}} = \text{Id}$ and channels where two of the p_i , with $i \in \{1, 2, 3\}$, are equal to zero. In the latter case, applying one of the Pauli matrices $\sigma_i \in \{\sigma_1, \sigma_2, \sigma_3\} = \{X, Y, Z\}$ with the same index i as the zero p_i to the channel before inputting in the superswitch maps the channel to a channel with only one zero p_i , thus achieving a nonzero probability $q_{---+}^{(2)}$ according to Result 1. We remark that among all unital channels 90.7% are antidegradable (see Appendix E), they have zero quantum capacity, and as a result, no protocol with definite causal order can ever distil them to the identity. Moreover, even with the quantum switch, only a set of measure zero among the latter can be distilled. Interestingly, in Appendix A we provide evidence that SWITCH_2 achieves a higher distillation rate than 4-switches that superpose all $4! = 24$ orderings of the four channels. Specifically, among the full 4-switches we considered, we found one instance of nonzero distillation rate equal to $\frac{4\mathcal{R}_{\mathcal{E}}^{(2)}}{9}$, a 44% decrease. We also considered restricted 4-switches that superpose only 16 orders and found a distillation rate of $\frac{3\mathcal{R}_{\mathcal{E}}^{(2)}}{4}$. We conjecture that among all possible restricted and unrestricted 4-switches, SWITCH_2 achieves the highest distillation rate.

Result 1 is a consequence of the fact that the channel corresponding to the outcome “ $---+$ ” for SWITCH_2 after measurement in the ancilla qubits is the identity channel, i.e., $\mathcal{E}_{---+}^{(2)} = \text{Id}$ (see Appendix B). The reason that this is possible for SWITCH_2 and not the quantum switch is due to the following two facts: (1) for any Pauli channel input in the quantum switch, a “ $-$ ” result on the ancilla results in a Pauli channel with $p_0 = 0$, and (2) for any Pauli channel with $p_0 = 0$ input in the quantum switch, a “ $+$ ” outcome leads to the identity channel. Thus, even though the quantum switch can lead to the identity channel for the input channels with $p_0 = 0$, or any of the other three faces of the tetrahedron by premultiplying with a unitary, it cannot lead to the identity for any channel in the interior: the quantum switch cannot distil any channel inside the tetrahedron. In contrast, as SWITCH_2 is a switch of switches, both (1) and (2) can simultaneously occur, leading to nonzero distillation in the interior.

Returning to the distillation rate in Result 1, we make the counterintuitive observation:

Result 2. Given two Pauli channels $\mathcal{E} = \mathcal{P}_{\vec{p}}$ and $\mathcal{F} = \mathcal{P}_{\vec{q}}$ such that $p_i \geq q_i$, $\forall i \in \{1, 2, 3\}$, we have $\mathcal{R}_{\mathcal{E}}^{(2)} \geq \mathcal{R}_{\mathcal{F}}^{(2)}$, with strict inequality if for at least one value of i , we have $p_i > q_i$. That is, the noisier channel $\mathcal{E}$ will have a higher distillation rate with the SWITCH₂ protocol.

The proof is given in Appendix D.

V. DISTILLATION PROTOCOL WITH SWITCH_n

For distillation protocols with SWITCH_n, a similar argument holds. As these are nested switches, any sequence of outcomes s that ends in “ $\dots - +$ ” implies that the channel obtained is the identity. To account for all contributions to the distillation rate in protocols with SWITCH_n, we can set up recurrence relations that can be iterated to obtain the distillation rates. We first describe the construction in the interior before examining the faces of the tetrahedron. At order n , we can write the channel of SWITCH_n as

$$\begin{aligned} \mathcal{S}^{(n)} = & \alpha^{(n)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n)} \mathcal{N}_{\vec{q}^{(n)}} \otimes \rho_{\mathcal{N}} \\ & + \gamma^{(n)} \mathcal{P}_{\vec{p}^{(n)}} \otimes \rho_{\mathcal{P}}, \end{aligned} \quad (6)$$

where $\rho_{\mathbf{Id}}$, $\rho_{\mathcal{N}}$, and $\rho_{\mathcal{P}}$ denote unnormalized states with orthogonal supports that gather the outcomes $o_{\mathbf{Id}}$, $o_{\mathcal{N}}$, and $o_{\mathcal{P}}$ of measurements on the ancillas that lead to the identity channel $\mathbf{Id}$, a channel with zero portion of the identity $\mathcal{N}$, i.e., $q_0^{(n)} = 0$, and a general Pauli channel with $p_0^{(n)} \neq 0, 1$, respectively. For instance, in the case of the second-order superswitch, we have $\rho_{\mathbf{Id}} = | - - + \rangle \langle - - + |$, $\rho_{\mathcal{N}} = | + + - \rangle \langle + + - | + | - + - \rangle \langle - + - | + | + - - \rangle \langle + - - | + | - - - \rangle \langle - - - |$, $\rho_{\mathcal{P}} = \mathbb{1} - \rho_{\mathbf{Id}} - \rho_{\mathcal{N}}$ (see Appendix J). The fact that it is possible to summarize all different channels into essentially three channels follows from the fact that for Pauli channels $\mathcal{P}_{\vec{p}_i}$ we have $\sum_i \lambda_i \mathcal{P}_{\vec{p}_i} = (\sum_i \lambda_i) \mathcal{P}_{\vec{p}'}$, with $\vec{p}' = \frac{\sum_i \lambda_i \vec{p}_i}{\sum_i \lambda_i}$.

Given that superswitches are switches of switches, and that the action of the superswitch at order n is assumed to take the form in Eq. (6), we can explicitly evaluate the superswitch at order $n+1$. In practice, this means that for each combination of terms we need to evaluate the contribution from the anticommutator and commutator from Eq. (3). There are nine possible combinations, each of which leads to two possible outcomes, so in principle we will have 18 terms at order $n+1$. We find nine unique terms, which can be rearranged and written in the form $\mathcal{S}^{(n+1)} = \alpha^{(n+1)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n+1)} \mathcal{N}_{\vec{q}^{(n+1)}} \otimes \rho_{\mathcal{N}} + \gamma^{(n+1)} \mathcal{P}_{\vec{p}^{(n+1)}} \otimes \rho_{\mathcal{P}}$. The equations for $\alpha^{(n+1)}$, $\beta^{(n+1)}$, $\gamma^{(n+1)}$, $\vec{q}^{(n+1)}$, and $\vec{p}^{(n+1)}$ define recurrence relations that can be iterated to obtain any order superswitch (see Appendix J). Even though we cannot solve the recurrence relations exactly, we look for their fixed points. We find exactly one solution: $\alpha^{(\infty)} = \frac{(2-\sqrt{3})}{4}$, $\beta^{(\infty)} = \frac{\sqrt{3}}{4}$, $\gamma^{(\infty)} = \frac{1}{2}$, $q_1^{(\infty)} = q_2^{(\infty)} = q_3^{(\infty)} = \frac{1}{3}$, and $p_1^{(\infty)} = p_2^{(\infty)} = p_3^{(\infty)} = \frac{(3-\sqrt{3})}{6}$. Thus, the asymptotic distillation rate is $\mathcal{R}_{\text{in}}^{(\infty)} = \frac{(2-\sqrt{3})}{4} \approx 0.067$. Note that the asymptotic distillation rate does not always imply the highest achievable distillation rate (see Appendix K).

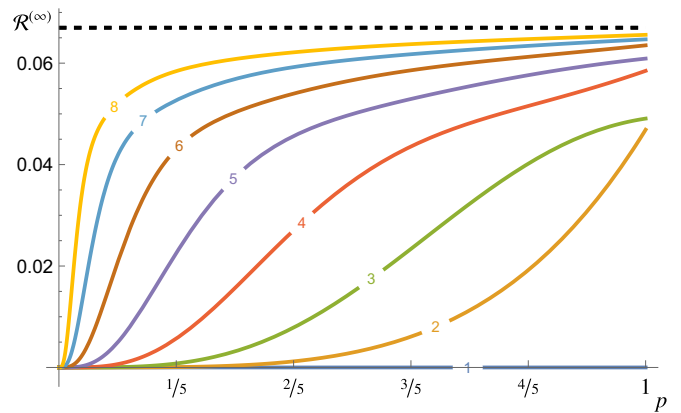


FIG. 3. The distillation rates $\mathcal{R}^{(n)}$ achieved by the superswitches for orders $n = 1, \dots, 8$. The quantum switch cannot distil for any value of p , while each superswitch achieves a higher distillation rate for all values of $p \in [0, 1]$, in comparison to all previous-order superswitches.

In the case of channels on the faces of the tetrahedron, unitary channels can also contribute to the distillation rate, which changes the form of the recurrence relation. We find that the asymptotic distillation rate on the base of the tetrahedron, $\mathcal{R}_{\text{base}}^{(\infty)}$ is different to the rate on the other three faces, $\mathcal{R}_{\text{in}}^{(\infty)}$ (see Appendix J). In summary, for the full set of Pauli channels, we have the following result:

Result 3. The asymptotic distillation rate inside the tetrahedron is $\mathcal{R}_{\text{in}}^{(\infty)} = \frac{(2-\sqrt{3})}{4} \approx 0.067$. On the three faces of channels with nonzero identity part, i.e., $p_0 \neq 0$, the distillation rate is $\mathcal{R}_{\text{in}}^{(\infty)} = \frac{(3-\sqrt{3})}{4} \approx 0.396$, while on the face of channels with $p_0 = 0$, it is $\mathcal{R}_{\text{base}}^{(\infty)} = \frac{1}{4}$.

VI. EXAMPLES

We now demonstrate our results for the depolarization channel, $\mathcal{D}(\rho) = (1 - \frac{3p}{4})\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z)$ with $p \in [0, \frac{4}{3}]$. Distillation is impossible through a depolarization channel with $p \neq \frac{4}{3}$ with the quantum switch. However, given access to SWITCH₂, distillation becomes possible for all values of p . Specifically, the distillation rate is $\mathcal{R}_{\text{dep}}^{(2)}(p) = \frac{3p^4}{64}$, which is nonzero for all values of p and, moreover, it is a monotonically increasing function of the channel noise parameter p , demonstrating Results 1 and 2. By iterating the recurrence relations (see Appendixes F, J, and K), the distillation rates of higher-order superswitches further increase the distillation rate for all values of $p \in [0, 1]$. In Fig. 3, we show a plot of the distillation rates for SWITCH_n with $n \leq 8$. The depolarization channel is entanglement breaking for all values of the noise parameter $p \in [\frac{2}{3}, \frac{4}{3}]$ and antidegradable for $p \in [\frac{1}{3}, \frac{4}{3}]$ (see Appendix E). Thus, for $p \geq \frac{1}{3}$, a depolarization channel has zero quantum capacity and no protocol with definite causal order can distil it to the identity. Interestingly, this is the range where the distillation rate is the highest for all orders n , echoing Result 2.

VII. EXTENSION TO NONUNITAL CHANNELS

So far, we focused our discussion on Pauli channels. However, the case of nonunital channels $\mathcal{N} = \sum_i K_i \rho K_i^\dagger$

can be covered by employing twirling [43,44]. Each copy of the channel can be turned into a depolarization channel with parameter $\eta = (\sum_i |\text{tr} K_i^2| - 1)/3$ and subsequently combined into the SWITCH₂ leading to the distillation rate $\mathcal{R}_{\mathcal{N}}^{(2)} = 3\eta^4/4$. While this comes at a great resource cost, channel-specific-twirling protocols may be able to reduce that overhead (see Appendix H).

VIII. NO-GO RESULT FOR MULTIQUBIT PAULI CHANNELS

In the preceding, we only considered the case of qubit channels. It is worth asking whether the above results can be generalized to higher dimensions. In Ref. [21], superswitches were also defined in any dimension 2^l with $l \geq 1$ and dimension $d = 4$ was discussed in detail. We have the following negative result.

Result 4. In any dimension $d = 2^l$ with $l > 1$, no superswitch of any order can lead to noiseless communication for all Pauli channels.

Thus, Results 1 and 2 no longer hold for higher dimensions. The proof is given in Appendix I. In summary, Result 1 is based on two facts: (1) for any channel, a minus result on the ancillas gives a channel with $p_0 = 0$, and (2) for any Pauli channel with $p_0 = 0$, a plus outcome leads to the identity channel. While (1) still holds, (2) fails. This follows from the commutators and anticommutators of tensor products of Pauli matrices.

IX. DISCUSSION

We have shown that exact channel distillation is possible probabilistically for any channel in dimension $d = 2$, including the set of channels with zero quantum capacity, such as entanglement-breaking and antidegradable channels. This proves that all channels can be used for perfect classical and quantum communication if two parties can implement at least the second-order superswitch and have access to a sufficient number of copies of the communicated states and channel. Owing to the Choi-Jamiołkowski isomorphism [39,40], access to higher-order superswitches thus implies that any channel in dimension $d = 2$ can be used to distribute entanglement with nonzero probability, including all channels of zero quantum capacity such as the set of entanglement-breaking and antidegradable channels (see Appendix E). Useless channels from the perspective of sharing entanglement [41], even those that remain useless after using the quantum switch, can thus be used to distribute entanglement given access to higher-order superswitches (see Appendix G). We have also found that, counterintuitively, the closer a channel is to the identity channel, the worse it performs at this task.

There are a number of interesting avenues for further research. Owing to the fact that the quantum switch has been implemented in practice in a number of different experiments, a direct adaptation of the setups to implement the second-order superswitch appears within reach, thus making our proposal for perfect quantum communication directly implementable with current technology.

Another interesting direction is to focus on dimensions $d > 2$ with the aim of designing switchlike supermaps that can achieve noise-free communication for all unital channels.

Finally, our results assume that a perfect implementation of the superswitches is possible, which, however, may not be realistic in practical implementations. A promising future direction is to revisit our results in the case where the ancilla qubits are assumed to interact with their environment, thus not being in perfect coherent superpositions. Quantifying the trade-off between coherence and fidelity under realistic noise models would enable analysis of our protocol in practical settings.

ACKNOWLEDGMENTS

S.K., J.M., and H.K. were supported by KIAS individual Grants No. CG086202 (S.K.), No. QP088702 (J.M.), and No. CG085302 (H.K.) at the Korea Institute for Advanced Study.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: COMPARISON OF SWITCH₂ TO THE 4-SWITCH

In this Appendix, we compare our distillation protocols, which are based on the superswitches SWITCH_n defined in the main text, to the versions of the n -switch [7,15,19]. The n -switch consists of all possible $n!$ orderings of n input channels and it is the most general object that one can define when controlling the ordering of n channels. The general form of n -switch is given by

$$n\mathcal{S}_\omega(\mathcal{E}, \mathcal{F}, \dots) = \sum_{i,j,\dots} W_{i,j,\dots}(\rho \otimes \omega) W_{i,j,\dots}^\dagger, \quad (\text{A1})$$

where the Kraus operators are defined as

$$W_{i,j,\dots} = \sum_{k=1}^{n!} \pi_k(E_i F_j \dots) \otimes |k\rangle\langle k|. \quad (\text{A2})$$

π_k denotes the k th permutation of the products of Kraus operators $E_i, F_j, \dots$ of the channels $\mathcal{E}, \mathcal{F}, \dots$ combined in the n -switch; $|k\rangle\langle k|_{k=1}^{n!}$ denotes an orthonormal basis in the control system. Both the state of the control ω and the measurement basis can be chosen freely. Their choice affects the resulting channels after each measurement outcome, as well as the probabilities of occurrence. In the conventional quantum switch, for example, the measurement basis is taken to be the Fourier basis (or equivalently the Hadamard basis in $d = 2$), which is mutually unbiased to the computational basis. In the case of the SWITCH_n, the choice is similar. Generalizing this approach, let H denote some normalized Hadamard matrix of order $n!$. Define the measurement basis $\{|h_k\rangle\langle h_k|\}_{k=1}^{n!}$ with

$$|h_k\rangle = H|k\rangle, \quad k = 1, \dots, n!, \quad (\text{A3})$$

and also choose the control system to be in the state $|h_1\rangle\langle h_1|$. As every Hadamard matrix can be cast in an equivalent form where its first row and column consist only of 1s, the control state is always $|h_1\rangle = \frac{1}{\sqrt{n!}} \sum_k |k\rangle$. It is this choice of the coherent superposition of all orders that allows the maximal

advantages through the switches. If, in the other extreme, the control was chosen to be one of the basis vectors $|k\rangle\langle k|$, we would only have the definite order corresponding to the permutation $\pi_k(E_i F_j \dots)$.

The resulting channel on the system multiplied by its probability of occurrence, $\tilde{\mathcal{E}}_r^{(n)}$, when measuring in the $\{|h_k\rangle\langle h_k|\}_{k=1}^{n!}$ basis and after obtaining measurement outcome $k = r$ is given by

$$\begin{aligned}\tilde{\mathcal{E}}_r^{(n)} &= \text{tr}[\mathbb{1} \otimes |h_r\rangle\langle h_r| n\mathcal{S}_\omega] \\ &= \sum_{i,j,\dots} \left[\left(\sum_k \pi_k(E_i F_j \dots) \overline{H_{kr}} H_{k1} \right) \rho \right. \\ &\quad \times \left. \left(\sum_k \pi_k(E_i F_j \dots) \overline{H_{kr}} H_{k1} \right)^\dagger \right].\end{aligned}\quad (\text{A4})$$

From the last result, we obtain the corresponding channel

$$\mathcal{E}_r^{(n)} = \frac{\tilde{\mathcal{E}}_r^{(n)}}{\text{Pr}(\tilde{\mathcal{E}}_r^{(n)})}, \quad (\text{A5})$$

where

$$\text{Pr}(\tilde{\mathcal{E}}_r^{(n)}) = \text{tr}(\tilde{\mathcal{E}}_r^{(n)}) \quad (\text{A6})$$

denotes the probability of the outcome r occurring.

We now restrict to the case $n = 4$ and compare SWITCH₂ to the full 4-switch. The question we ask is whether the full 4-switch can be used for distillation, and if so, what is the distillation rate?

We recall that for any Pauli channel $\mathcal{P}_{\vec{p}}$, four copies of which are combined in the second-order superswitch, SWITCH₂, there is a nonzero probability that the outcome “− − +” that leads to the identity channel will be obtained. Specifically, we found that the distillation rate $\mathcal{R}^{(2)}$ is

$$\mathcal{R}^{(2)} = 4(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2). \quad (\text{A7})$$

To compare with the full 4-switch, we choose a number of known Hadamard matrices and apply Eq. (A4). In the case $n = 4$, we have to choose a Hadamard matrix of order $m = 4! = 24$, the number of permutations of the ordering in the product of the four Kraus operators. We note that for every choice H , its transpose $H^\top$ is also a legitimate choice, as both rows and columns of a Hadamard matrix form orthonormal bases that are mutually unbiased to the computational one.

Following the intuition from the case of the quantum switch, the obvious choice is to take the Fourier matrix with elements

$$F_{j,k}^{(24)} = \frac{1}{\sqrt{24}} e^{\frac{i\pi}{12}(j-1)(k-1)}, \quad j, k \in \{1, \dots, 24\}. \quad (\text{A8})$$

However, we find that none of the 24 channels is the identity channel. As a result, distillation is not possible with the Fourier matrix, and we arrive at the same conclusion for its transpose. Next, we consider the case of real Hadamard matrices. Real Hadamard matrices of order $m = 24$ have been completely classified [45,46], and it is known that the number of inequivalent real Hadamard matrices is 60. We check all 60 matrices, as well as their transposes, and we find that only a

single instance gives the identity channel. This happens for the transpose of the matrix number 27 in Ref. [45] after a measurement on the control that yields outcome $r = 18$. The distillation rate in this case is given by the probability of this outcome occurring. We find that it differs by a constant factor from the distillation rate of SWITCH₂,

$$\mathcal{R}^{(4,27)} = \frac{4}{9} \mathcal{R}^{(2)} = \frac{16(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2)}{9}. \quad (\text{A9})$$

Thus, we see that $\mathcal{R}^{(4,27)} < \mathcal{R}^{(2)}$.

This is the only case we found where distillation happens for the 4-switch. Owing to the fact that a classification of complex Hadamard matrices in dimension 24 is an open problem, we are not able to answer the question in full generality; however, our preliminary analysis suggests that restricting to a subspace of the 4-switch may be beneficial for the distillation task we considered. Indeed, the SWITCH₂ is obtained from the 4-switch by restricting to an eight-dimensional subspace. Essentially, from the set of all 24 orders, SWITCH₂ applies only 8 possible orders. Given that it can be analytically shown that the algebraic structure $\sum_{i,j,k,l} \{[E_i, E_j], [E_k, E_l]\} \rho \{[E_i, E_j], [E_k, E_l]\}^\dagger \propto \rho$ for any Pauli channel and this requires the superposing of eight different orders to materialize as a generalized switch, we do not expect restricting to a subspace with $d < 8$ to lead to the identity channel for all Pauli channels. To that end, we also examined the case of the 3-switch for the Fourier matrix (no real Hadamard matrices of order $3! = 6$ exist) and found that distillation is not possible.

To conclude, we note that the term of the real Hadamard matrix number 27 that corresponds to measurement outcome $r = 18$ takes the form

$$\begin{aligned}\mathcal{E}_{18}^{(4,27)} &= \frac{3}{24^2} \sum_{i,j,k,l} \{[E_i, E_j], [E_k, E_l]\} \rho \{[E_i, E_j], [E_k, E_l]\}^\dagger \\ &\quad + \text{cross terms.}\end{aligned}\quad (\text{A10})$$

Interestingly, this confirms the intuition derived from SWITCH₂, namely, that the algebraic structure related to the nesting of the anticommutator with two inside commutators of Kraus operators is what makes the distillation possible. However, in this case it also follows that the cross terms lead to destructive interference, reducing the distillation rate by a factor $\frac{4}{9} \approx 0.44$. We also checked the case of keeping only two instead of three terms in the brackets in Eq. (A10) corresponding to restricting to a 16-dimensional subspace, and the guessing probability increases to $\frac{3\mathcal{R}^{(2)}}{4}$, which provides further evidence that restricting to a subspace of the full 4-switch is beneficial for distillation. We conjecture that the protocol with SWITCH₂ achieves the best distillation rate among all possible protocols with the 4-switch but we defer a more thorough investigation to future work.

APPENDIX B: n -ORDER SUPERSWITCHES

The n -order superswitch can be iteratively defined as a switch of two $(n-1)$ -order switches [21]. By denoting the Kraus operators of the n -order superswitch as $K^{(n)}$, these are

explicitly defined in terms of those of the previous orders as

$$\begin{aligned} K_{\mathbf{i}_{n-1}\mathbf{i}'_{n-1}}^{(n)} &= K_{\mathbf{i}_{n-1}}^{(n-1)} \tilde{K}_{\mathbf{i}'_{n-1}}^{(n-1)} \otimes |0\rangle\langle 0|_{2^{n-1}} + \tilde{K}_{\mathbf{i}'_{n-1}}^{(n-1)} K_{\mathbf{i}_{n-1}}^{(0)} \otimes |1\rangle \\ &\times \langle 1|_{2^{n-1}} \\ &= \frac{1}{2} \{K_{\mathbf{i}_{n-1}}^{(n-1)}, \tilde{K}_{\mathbf{i}'_{n-1}}^{(n-1)}\} \otimes \mathbb{1} + \frac{1}{2} [K_{\mathbf{i}_{n-1}}^{(n-1)}, \tilde{K}_{\mathbf{i}'_{n-1}}^{(n-1)}] \otimes Z, \end{aligned} \quad (\text{B1})$$

where the index of the control qubit denotes that at order n there are $2^n - 1$ control qubits in total. The index

symbol in bold, $\mathbf{i}_{n-1}$, is shorthand for the list of indices $\mathbf{i}_{n-1} \equiv i_1 i_2 \cdots i_{2^{n-1}}$ associated with the Kraus operators of the superswitch of order $n - 1$. We use the convention $K^{(0)} = E_i, \tilde{K}^{(0)} = F_j, \dots$; that is, at order zero we only have the Kraus operators of the channels to be combined in a superswitch. By expanding the nested expressions that consist of all commutators and anticommutators of all pairs of commutators and anticommutators generated by the Kraus operators of the channels, we get the following expression for SWITCH_n :

$$\begin{aligned} \mathcal{S}_{\omega_1 \cdots \omega_{2^n-1}}^{(n)}(\mathcal{E}, \mathcal{F}, \dots) &= \frac{1}{2^{2^n-1-2}} \sum_{\substack{i_1, \dots, i_{2^n-1} = \pm \\ i, j, \dots = 1, \dots, 4}} ([\dots, [[E_i, F_j]_{i_1}, [\tilde{E}_k, \tilde{F}_l]_{i_2}], \dots]_{i_{2^n-1}} \rho [\dots, [[E_i, F_j]_{i_1}, [\tilde{E}_k, \tilde{F}_l]_{i_2}], \dots]_{i_{2^n-1}} \\ &\otimes P_{i_1} \omega_1 P_{i_1} \otimes \cdots \otimes P_{i_{2^n-1}} \omega_{2^n-1} P_{i_{2^n-1}}), \end{aligned} \quad (\text{B2})$$

with $P_+ = \mathbb{1}$ and $P_- = Z$. Explicit expressions for a given order n can be evaluated iteratively [21].

In this framework, the first-order superswitch is the quantum switch, which given two channels $\mathcal{E}$ and $\mathcal{F}$ with Kraus operators E_i and F_j , respectively, is

$$\begin{aligned} \mathcal{S}_{\omega}^{(1)}(\mathcal{E}, \mathcal{F})(\rho) &= \sum_{i,j} K_{ij}^{(1)} \rho \otimes \omega K_{ij}^{(1)} \\ &= \frac{1}{4} \sum_{i,j} \{E_i, F_j\} \rho \{E_i, F_j\}^\dagger \otimes \omega \\ &+ \frac{1}{4} \sum_{i,j} [E_i, F_j] \rho [E_i, F_j]^\dagger \otimes Z \omega Z, \end{aligned} \quad (\text{B3})$$

matching the expression in the main text [Eq. (2)]. The last expression shows that the quantum switch effectively evaluates channels that follow from commutator and anticommutators of the Kraus operators of the channels. Nontrivial effects occur whenever some of the Kraus operators do not commute with each other.

We now explicitly derive the second-order superswitch, which is crucial for the results in the main text. Specifically, with four channels $\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}},$ and $\tilde{\mathcal{F}}$ that have Kraus operators

$E_i, F_j, \tilde{E}_k,$ and $\tilde{F}_l$, respectively, we initially define two ordinary quantum switches, that is, two first-order superswitches. Their Kraus operators, $K_{ij}^{(1)}$ and $\tilde{K}_{ij}^{(1)}$, are explicitly

$$\begin{aligned} K_{ij}^{(1)} &= E_i F_j \otimes |0\rangle\langle 0|_1 + F_j E_i \otimes |1\rangle\langle 1|_1, \\ \tilde{K}_{ij}^{(1)} &= \tilde{E}_i \tilde{F}_j \otimes |0\rangle\langle 0|_2 + \tilde{F}_j \tilde{E}_i \otimes |1\rangle\langle 1|_2, \end{aligned} \quad (\text{B4})$$

with the subscripts differentiating between the different ancilla qubits that control the ordering of the channels $\mathcal{E}, \mathcal{F}$ and $\tilde{\mathcal{E}}, \tilde{\mathcal{F}}$, respectively. The Kraus operators of the second-order superswitch are then defined as

$$\begin{aligned} K_{ijkl}^{(2)} &= K_{ij}^{(1)} \tilde{K}_{kl}^{(1)} \otimes |0\rangle\langle 0|_3 + \tilde{K}_{kl}^{(1)} K_{ij}^{(1)} \otimes |1\rangle\langle 1|_3 \\ &= \frac{1}{2} \{K_{ij}^{(1)}, \tilde{K}_{kl}^{(1)}\} \otimes \mathbb{1} + \frac{1}{2} [\tilde{K}_{kl}^{(1)}, K_{ij}^{(1)}] \otimes Z. \end{aligned} \quad (\text{B5})$$

We readily obtain the second-order superswitch as

$$\mathcal{S}_{\omega_c}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) = \sum_{i,j,k,l} K_{ijkl}^{(2)} (\rho \otimes \omega_c) (K_{ijkl}^{(2)})^\dagger, \quad (\text{B6})$$

where the ancilla state ω_c controls the order of channels $\mathcal{E}, \mathcal{F}$ and $\tilde{\mathcal{E}}, \tilde{\mathcal{F}}$ inside the two first-order superswitches, respectively, as well as the ordering of the two superswitches themselves. A schematic representation is shown in Fig. 4. Expanding the Kraus operators, we find

$$\begin{aligned} K_{ijkl}^{(2)} &= E_i F_j \tilde{E}_k \tilde{F}_l \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| + E_i F_j \tilde{F}_l \tilde{E}_k \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| + F_j E_i \tilde{E}_k \tilde{F}_l \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| \\ &+ F_j E_i \tilde{F}_l \tilde{E}_k \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| + \tilde{E}_k \tilde{F}_l E_i F_j \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1| + \tilde{F}_l \tilde{E}_k E_i F_j \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1| \\ &+ \tilde{E}_k \tilde{F}_l F_j E_i \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1| + \tilde{F}_l \tilde{E}_k F_j E_i \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|, \end{aligned} \quad (\text{B7})$$

where we have dropped the subscripts from the ancilla qubits. This expression can also be written in terms of commutators and anticommutators as

$$\begin{aligned} K_{ijkl}^{(2)} &= \frac{1}{8} \sum_{s_1, s_2, s} [[E_i, F_j]_{s_1}, [\tilde{E}_k, \tilde{F}_l]_{s_2}]_s \otimes P_{s_1} \otimes P_{s_2} \otimes P_s \frac{1}{8} (\{[E_i, F_j], \{\tilde{E}_k, \tilde{F}_l\}\} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} + \{[E_i, F_j], \{\tilde{E}_k, \tilde{F}_l\}\} \otimes Z \otimes \mathbb{1} \otimes \mathbb{1} \\ &+ \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \otimes \mathbb{1} \otimes Z \otimes \mathbb{1} + \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \otimes Z \otimes Z \otimes \mathbb{1} + \{[E_i, F_j], \{\tilde{E}_k, \tilde{F}_l\}\} \otimes \mathbb{1} \otimes \mathbb{1} \otimes Z \\ &+ [[E_i, F_j], \{\tilde{E}_k, \tilde{F}_l\}] \otimes Z \otimes \mathbb{1} \otimes Z + \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \otimes \mathbb{1} \otimes Z \otimes Z + [[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]] \otimes Z \otimes Z \otimes Z), \end{aligned} \quad (\text{B8})$$

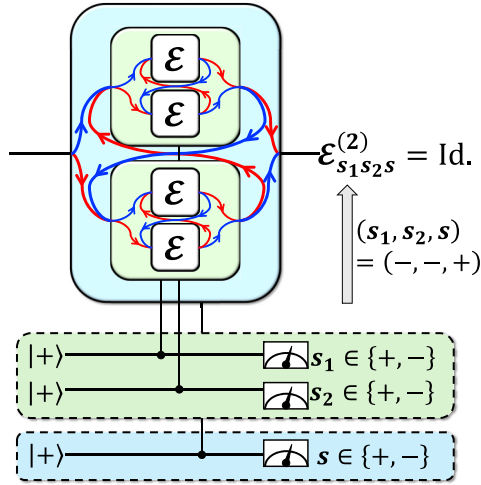


FIG. 4. A visual representation of the second-order superswitch.

where we have introduced the notation $[A, B]_+ \equiv [A, B] = AB + BA$ and $[A, B]_- \equiv [A, B] = AB - BA$ for the anticommutator and commutator, and $P_+ = \mathbb{1}$, $P_- = Z$. We now assume that for any pair of indices i and j , the commutator $[X_i, Y_j]$ and anticommutator $\{X_i, Y_j\}$ of a pair of Kraus operators X_i, Y_j are not simultaneously nonzero, which holds for Pauli channels; we then obtain the second-order superswitch,

$$\begin{aligned} \mathcal{S}_{\omega_c}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) &= \sum_{i,j,k,l} K_{ijkl}(\rho \otimes \omega_c) K_{ijkl}^\dagger \\ &= \frac{1}{64} \sum_{i,j,k,l} \sum_{s_1, s_2, s} [[E_i, F_j]_{s_1}, [\tilde{E}_k, \tilde{F}_l]_{s_2}]_s \\ &\quad \times \rho [[E_i, F_j]_{s_1}, [\tilde{E}_k, \tilde{F}_l]_{s_2}]_s^\dagger \\ &\quad \otimes (P_{s_1} \otimes P_{s_2} \otimes P_s) \omega_c (P_{s_1} \otimes P_{s_2} \otimes P_s)^\dagger, \end{aligned} \quad (\text{B9})$$

where the indices $i, j, k, l \in [1, \dots, 4]$ and $s_1, s_2, s = \pm$.

We further choose the ancilla state to be a product state $\omega_c = \omega_1 \otimes \omega_2 \otimes \omega$, with ω_1 and ω_2 controlling the ordering of the channels $\mathcal{E}, \mathcal{F}$ and $\tilde{\mathcal{E}}, \tilde{\mathcal{F}}$, respectively, in the first-order superswitches and ω controlling their ordering. Letting $\omega_1 = \omega_2 = \omega = |+\rangle\langle+|$ and defining channels $\mathcal{E}_{ijk}^{(2)}$ as, e.g.,

$$\begin{aligned} \mathcal{E}_{--+}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) &= \frac{1}{q_{--+}} \frac{1}{64} \sum_{i,j,k,l} \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \rho \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\}^\dagger, \end{aligned} \quad (\text{B10})$$

where

$$\begin{aligned} q_{--+}^{(2)} &= \text{tr} \left(\frac{1}{64} \sum_{i,j,k,l} \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \rho \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\}^\dagger \right) \end{aligned} \quad (\text{B11})$$

are the associated probabilities of occurrence, the second-order superswitch takes the form

$$\begin{aligned} \mathcal{S}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) &= \sum_{s_1, s_2, s = \pm} q_{s_1 s_2 s}^{(2)} \mathcal{E}_{s_1 s_2 s}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}}) \otimes |s_1 s_2 s\rangle \langle s_1 s_2 s| \\ &= q_{+++}^{(2)} \mathcal{E}_{+++}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |+++ \rangle \langle +++| \\ &\quad + q_{-++}^{(2)} \mathcal{E}_{-++}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |-++ \rangle \langle -++| \\ &\quad + q_{+-+}^{(2)} \mathcal{E}_{+-+}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |+-+ \rangle \langle +-+| \\ &\quad + q_{--+}^{(2)} \mathcal{E}_{--+}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |--+ \rangle \langle |--+| \\ &\quad + q_{++-}^{(2)} \mathcal{E}_{++-}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |++- \rangle \langle ++-| \\ &\quad + q_{+-+}^{(2)} \mathcal{E}_{+-+}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |+-+ \rangle \langle +-+| \\ &\quad + q_{+--}^{(2)} \mathcal{E}_{+--}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |+-- \rangle \langle +--| \\ &\quad + q_{-+-}^{(2)} \mathcal{E}_{-+-}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |-+- \rangle \langle |-+-| \\ &\quad + q_{---}^{(2)} \mathcal{E}_{---}^{(2)}(\mathcal{E}, \mathcal{F}, \tilde{\mathcal{E}}, \tilde{\mathcal{F}})(\rho) \otimes |--+ \rangle \langle |--+|. \end{aligned} \quad (\text{B12})$$

We let $\mathcal{E} = \tilde{\mathcal{E}} = \mathcal{F} = \tilde{\mathcal{F}} = \mathcal{P}_{\vec{p}} = \sum_i p_i \sigma_i \rho \sigma_i^\dagger$; that is, instances of the same channel are combined in the first-order superswitch, and we list the general expressions. For notational brevity, we will denote a Pauli channel $\mathcal{E} = \mathcal{P}_{\vec{p}} = p_0 \rho + p_1 X \rho X + p_2 Y \rho Y + p_3 Z \rho Z$ with the last three components of its probability vector $\vec{p}$, since, due to the normalization condition, the first follows from the other three, i.e., $p_0 = 1 - p_1 - p_2 - p_3$, and we will write $\mathcal{E} = \{p_1, p_2, p_3\}$. For instance, we will write $\mathcal{D}_p = \{\frac{p}{4}, \frac{p}{4}, \frac{p}{4}\}$ to denote a depolarization channel. With this notation, we can list all the tuples $\Omega_{ijk}^{(2)} = \{q_{ijk}^{(2)}, \mathcal{E}_{ijk}^{(2)}\}$, consisting of the eight $\mathcal{E}_{ijk}^{(2)}$ channels of the second-order superswitch:

$$\begin{aligned} \mathcal{E}_{+++}^{(2)} &= \frac{4p_0(p_0^2 + p_1^2 + p_2^2 + p_3^2)}{q_{+++}^{(2)}} \{p_1, p_2, p_3\}, \\ \mathcal{E}_{-++}^{(2)} &= \frac{2(p_0^2 + p_1^2 + p_2^2 + p_3^2)}{q_{-++}^{(2)}} \{p_2 p_3, p_1 p_3, p_1 p_2\}, \\ \mathcal{E}_{+-+}^{(2)} &= \mathcal{E}_{-++}^{(2)}, \\ \mathcal{E}_{--+}^{(2)} &= \{0, 0, 0\} = \text{Id}, \\ \mathcal{E}_{++-}^{(2)} &= \frac{8p_0^2}{q_{++-}^{(2)}} \{p_2 p_3, p_1 p_3, p_1 p_2\}, \\ \mathcal{E}_{-+-}^{(2)} &= \frac{4p_0}{q_{-+-}^{(2)}} \{p_1(p_2^2 + p_3^2), p_2(p_1^2 + p_3^2), p_3(p_1^2 + p_2^2)\}, \\ \mathcal{E}_{+--}^{(2)} &= \mathcal{E}_{-+-}^{(2)}, \\ \mathcal{E}_{---}^{(2)} &= \frac{8p_1 p_2 p_3}{q_{---}^{(2)}} \{p_1, p_2, p_3\}, \end{aligned} \quad (\text{B13})$$

with the respective probabilities of occurrence:

$$\begin{aligned} q_{+++}^{(2)} &= p_0^4 + 4p_0^3(p_1 + p_2 + p_3) + 6p_0^2(p_1^2 + p_2^2 + p_3^2) \\ &\quad + 4p_0(p_1 + p_2 + p_3)(p_1^2 + p_2^2 + p_3^2) \\ &\quad + (p_1^2 + p_2^2 + p_3^2)^2, \end{aligned}$$

$$\begin{aligned}
q_{--+}^{(2)} &= 12p_0p_1p_2p_3 \\
&\quad + 2(p_0^2 + p_1^2 + p_2^2 + p_3^2)(p_1p_2 + p_2p_3 + p_3p_1), \\
q_{+-+}^{(2)} &= q_{-++}^{(2)}, \\
q_{--+}^{(2)} &= 4(p_1^2p_2^2 + p_2^2p_3^2 + p_3^2p_1^2), \\
q_{++-}^{(2)} &= 8p_0^2(p_1p_2 + p_2p_3 + p_3p_1), \\
q_{-+-}^{(2)} &= 4p_0(p_1^2(p_2 + p_3) + p_2^2(p_3 + p_1) + p_3^2(p_1 + p_2)), \\
q_{+--}^{(2)} &= q_{-+-}^{(2)}, \\
q_{---}^{(2)} &= 8p_1p_2p_3(p_1 + p_2 + p_3). \tag{B14}
\end{aligned}$$

Note that the $\mathcal{E}_{--+}^{(2)}$ channel is the identity channel $\mathbf{Id}$, which is the crucial feature that enables our results.

We remark that in Ref. [20] the concept of switch of switches was introduced, similar to the case we consider here but in a restricted scenario where there is only a single control ancilla for the inside switches; i.e., their ordering is “synchronized.” Here, we use the general case that is crucial for our purposes, as the synchronized switches in Ref. [20] do not exhibit the same properties. The reason for this is that the effect of forcing synchronization leads to mixing of terms [21], and thus the possibility to get the identity channel is in general removed.

APPENDIX C: GEOMETRIC PICTURE OF PAULI CHANNELS IN $d = 2$

A completely positive and trace-preserving (CPTP) map acting on states in dimension $d = 2$ can be written in the general form

$$\tilde{\mathcal{E}}(\rho) = \mathcal{U} \circ \mathcal{E}(\rho) \circ \mathcal{V}, \tag{C1}$$

where $\mathcal{E}$ can be seen as acting on the Bloch vector of the state as

$$\mathcal{E}(\rho) = \frac{1}{2}(\mathbb{1} + (M\vec{r} + \vec{t}) \cdot \vec{\sigma}), \tag{C2}$$

with $M = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ and $\vec{t} = (t_1, t_2, t_3)$. λ_i satisfy $(\lambda_1 \pm \lambda_2)^2 \leq (1 \pm \lambda_3)^2$, along with permutations of the indices. The conditions on the elements of M and $\vec{t}$ for a map to be CPTP are given in Ref. [47]. Geometrically, a channel will map the Bloch ball into a displaced ellipsoid inside it. However, not all ellipsoids inside the Bloch ball correspond to CPTP maps.

The case $\vec{t} = 0$ corresponds to unital channels and in terms of the parametrization we have been using so far, i.e., $\mathcal{E}(\rho) = \sum_i p_i \sigma_i \rho \sigma_i$, we have

$$\lambda_i = 1 - 2p_j - 2p_k, \tag{C3}$$

with $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$.

The Kraus rank of a channel is the minimum number of Kraus operators among all Kraus decompositions. Kraus rank 1 maps are the $\mathbf{Id}$ channel and the three unitary channels $\mathcal{U}_X, \mathcal{U}_Y$, and $\mathcal{U}_Z$, which correspond to the vertices of the tetrahedron. Similarly, Kraus 2 operations are on the edges, while Kraus 3 are on the faces. Any map inside the tetrahedron necessarily has Kraus rank 4.

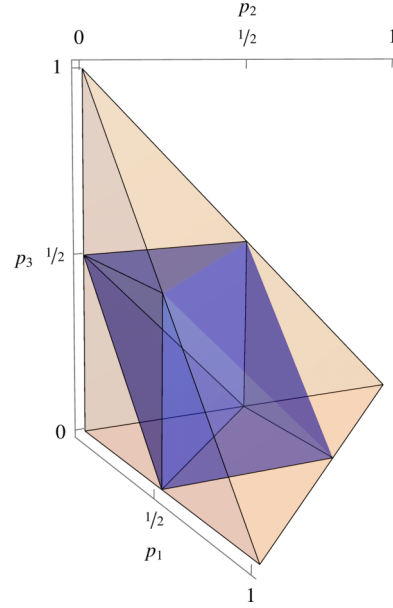


FIG. 5. The set of unital channels, corresponding to the tetrahedron in yellow. The octahedron in blue corresponds to the set of entanglement-breaking channels.

A Pauli channel is entanglement breaking if the following conditions are satisfied [48,49]:

$$\begin{aligned}
\lambda_1 + \lambda_2 + \lambda_3 &\leq 1, \quad \lambda_i - \lambda_j - \lambda_k \leq 1, \\
\forall i \neq j \neq k. \tag{C4}
\end{aligned}$$

It is straightforward to check that in terms of the p_i parametrization of unital channels, these conditions are equivalent to $p_i \leq \frac{1}{2}, \forall i$. These define an octahedron inside the tetrahedron of unital channels.

From the discussion in the main text, including Result 1, at the faces of the tetrahedron lie channels that can be purified by employing the quantum switch. Channels of Kraus rank 4 lie strictly inside the tetrahedron and can only be purified with the second-order superswitch. Thus, with the quantum switch only a particular subset of entanglement-breaking channels (i.e., those at the intersection of the surfaces of octahedron and tetrahedron) can be distilled. On the other hand, the second-order superswitch enables distillation of all entanglement-breaking channels. The tetrahedron of unital channels, as well as the octahedron corresponding to the set of entanglement-breaking channels, is shown in Fig. 5.

APPENDIX D: PROOF OF RESULTS 1 AND 2

Result 1 follows by noting that in the expressions for the second-order superswitch [Eqs. (B13) and (B14)], the outcome “— — +” after measurements in the ancillas (1) leads to the identity channel $\mathbf{Id}$, for *any* input Pauli channel $\mathcal{E}_{\vec{p}}$, and (2) this probability is nonzero, i.e., $q_{--+} > 0$, for all channels except for the trivial case of $\mathcal{E}_{\vec{p}} = \mathbf{Id}$ as well as channels with two of the p_i with $i = 1, 2, 3$ equal to zero. However, the latter case can be mapped to the case with only one of the p_i being zero. Specifically, applying one of the Pauli matrices σ_i with the same index i as the zero p_i to the channel before inputting in the superswitch maps each channel to a channel

with only one zero p_i , thus achieving nonzero probability $q_{--+}^{(2)}$ according to Result 1, and thus all channels are covered. For example, applying either a unitary Y or Z to the channel $(1-p)\rho + pX\rho X$ maps it to either $(1-p)Y\rho Y + pZ\rho Z$ or $(1-p)Z\rho Z + pY\rho Y$, respectively. Thus, all channels are covered by Result 1.

The fact that the second-order superswitch is a switch of two switches is what enables this phenomenon. Specifically, this follows directly from the equations of the quantum switch [4,23], in the case of two channels $\mathcal{E} = \mathcal{F} = p_0\rho + p_1X\rho X + p_2Y\rho Y + p_3Z\rho Z$ with $\sum_i p_i = 1$, which reduce to $S_\omega(\rho) = q_+C_+(\rho) \otimes \omega_+ + q_-C_-(\rho) \otimes \omega_-$. The channels C_+ and C_- [23] are defined as

$$C_+(\rho) = \frac{1}{q_+} \left(\left(\sum_{i=0}^3 p_i^2 \right) \rho + 2p_0 \left(\sum_{i=1}^3 p_i \sigma_i \rho \sigma_i \right) \right),$$

$$C_-(\rho) = \frac{2}{q_-} (p_1 p_2 \sigma_3 \rho \sigma_3 + p_2 p_3 \sigma_1 \rho \sigma_1 + p_3 p_1 \sigma_2 \rho \sigma_2), \quad (\text{D1})$$

with $q_- = 2(p_1 p_2 + p_2 p_3 + p_3 p_1)$, $q_+ = 1 - q_-$, $\omega_+ = \omega$, and $\omega_- = Z\omega Z$. The choice $\omega = |+\rangle\langle +|$ leads to $\omega_\pm = |\pm\rangle\langle \pm|$ and thus by measuring the ancilla on the same basis, the two channels can be fully separated.

It readily follows that (1) for any Pauli channel, two copies of which are combined in the quantum switch, the minus outcome after measurement on the ancilla qubit leads to a Pauli channel with $p_0 = 0$, and (2) for any Pauli channel with $p_0 = 0$, two copies of which are combined in the quantum switch, the plus outcome after measurement on the ancilla qubit leads to the identity channel Id . Then, a way to satisfy both conditions at the same time is to construct the second-order superswitch: whenever measurements on the ancillas of the two inside switches give minus outcomes and measurement on the ancilla controlling the order of the switches gives a plus outcome, the above scenario occurs and we thus obtain a noiseless copy.

Result 2 follows easily from the functional form of the distillation rate,

$$\mathcal{R}^{(2)} = q_{--+}^{(2)} = 4(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2), \quad (\text{D2})$$

and the observation that as all terms are positive, given two channels $\mathcal{E} = \mathcal{P}_{\vec{p}}$ and $\mathcal{F} = \mathcal{P}_{\vec{q}}$ with probability vectors $\vec{p}, \vec{q}$ such that $p_i \geq q_i$ for all $i \in \{1, 2, 3\}$, then we necessarily have $\mathcal{R}_{\mathcal{E}}^{(2)} \geq \mathcal{R}_{\mathcal{F}}^{(2)}$. The fact that $p_i \geq q_i$ for all $i \in \{1, 2, 3\}$ implies that the channel $\mathcal{E}$ is farther away from the identity channel Id .

To make this more mathematically precise, we compare the difference of their trace distances by defining the quantity

$$d(\mathcal{E}, \mathcal{F}, \rho) = T(\rho, \mathcal{E}(\rho)) - T(\rho, \mathcal{F}(\rho)), \quad (\text{D3})$$

where T denotes the trace distance,

$$T(\rho, \sigma) = \frac{\|\rho - \sigma\|_1}{2} = \frac{1}{2} \text{tr}(\sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)}). \quad (\text{D4})$$

Note that the quantity $d(\mathcal{E}, \mathcal{F}, \rho)$ is not a distance as it can be negative and is not symmetric in the channels. As for two qubit states ρ and σ , the trace distance is related to the Euclidean distance d_e of their Bloch vectors $\vec{r}$ and $\vec{s}$, respectively,

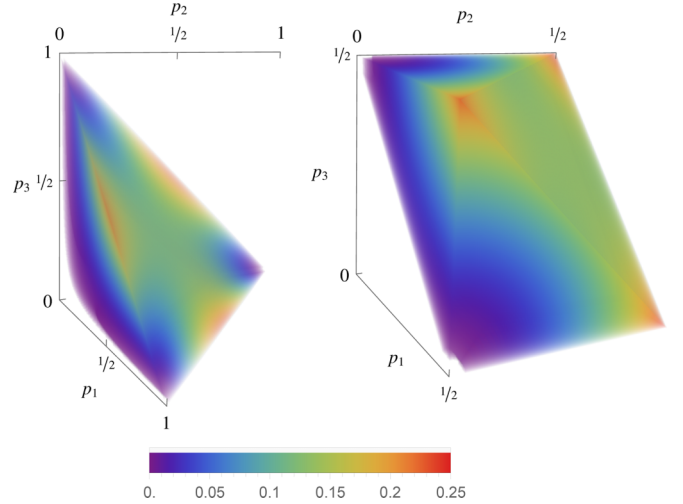


FIG. 6. Visualization of the distillation rate $\mathcal{R}^{(2)}$ for (1) all unital channels (left) and (2) the set of entanglement-breaking channels (right).

and we have

$$T(\rho, \sigma) = \frac{d_e(\vec{r}, \vec{s})}{2} = \frac{\sqrt{(r_1 - s_1)^2 + (r_2 - s_2)^2 + (r_3 - s_3)^2}}{2}. \quad (\text{D5})$$

The action of a quantum channel $\mathcal{N}$ in dimension $d = 2$ is [47]

$$\mathcal{N}(\rho) = \frac{1}{2}(\mathbb{1} + (M\vec{r} + \vec{t}) \cdot \vec{\sigma}). \quad (\text{D6})$$

Restricting to the case of a Pauli channel, the Bloch vector is transformed as

$$(r_1, r_2, r_3) \rightarrow ((1 - 2p_2 - 2p_3)r_1, (1 - 2p_3 - 2p_1)r_2, (1 - 2p_1 - 2p_2)r_3). \quad (\text{D7})$$

Then, for two Pauli channels $\mathcal{E} = \mathcal{P}_{\vec{p}}$ and $\mathcal{F} = \mathcal{P}_{\vec{q}}$ such that $p_i \geq q_i$ for all $i \in \{1, 2, 3\}$, we have

$$d(\mathcal{E}, \mathcal{F}, \rho) = \sqrt{r_1^2(p_2 + p_3)^2 + r_2^2(p_3 + p_1)^2 + r_3^2(p_1 + p_2)^2} - \sqrt{r_1^2(q_2 + q_3)^2 + r_2^2(q_3 + q_1)^2 + r_3^2(q_1 + q_2)^2} \geq 0, \quad (\text{D8})$$

which can be restated as the statement that $\forall \rho$

$$\|\rho - \mathcal{E}(\rho)\|_1 \geq \|\rho - \mathcal{F}(\rho)\|_1, \quad (\text{D9})$$

which proves the claim. The distillation rates for Pauli channels in the interior of the tetrahedron are visualized in Fig. 6.

APPENDIX E: COMPARISON TO PROTOCOLS WITH DEFINITE CAUSAL ORDER

We compare our distillation rates to protocols with definite causal order. It is clear that with operations in definite causal order, a zero-capacity channel can never be distilled to the identity or a unitary channel by any protocol.

Entanglement-breaking channels are the primal example of zero-capacity channels. Here, we evaluate another set of zero-capacity channels, the *antidegradable* channels, and show that they are a superset of entanglement-breaking channels and in fact cover 90.7% of the whole set of unital channels.

We briefly introduce degradable and antidegradable channels. Given a channel $\mathcal{E}$, it follows from the Stinespring dilation theorem that it can be written as

$$\mathcal{E}(\rho) = \text{tr}_E[U(\rho_S \otimes |0\rangle\langle 0|_E)U_{SE}^\dagger]. \quad (\text{E1})$$

Indices S and E denote the system and an environment. The *complementary channel* $\tilde{\mathcal{E}}$ of $\mathcal{E}$ is obtained by tracing out the system instead of the environment, that is,

$$\tilde{\mathcal{E}}(\rho) = \text{tr}_S[U_{SE}(\rho_S \otimes |0\rangle\langle 0|_E)U_{SE}^\dagger]. \quad (\text{E2})$$

A channel is called *degradable* if there exists another channel $\mathcal{F}$ such that

$$\tilde{\mathcal{E}} = \mathcal{F} \circ \mathcal{E}; \quad (\text{E3})$$

that is, the complementary channel can be simulated from the system's output by applying the channel $\mathcal{F}$. Similarly, a channel is *antidegradable* if its complementary channel is degradable. That is, there is a channel $\mathcal{F}$ such that

$$\mathcal{E} = \mathcal{F} \circ \tilde{\mathcal{E}}. \quad (\text{E4})$$

What is crucial for our purposes is that antidegradable channels have zero quantum capacity. The case of antidegradable unital qubit channels has been studied in Refs. [50–52], and a necessary and sufficient condition for a channel to be antidegradable has been derived.

Specifically, let $m \in \{1, 2, 3, 4\}$ the index that corresponds to the probability p_i in the definition of the Pauli channel, $\mathcal{E} = \sum_i p_i \sigma_i \rho \sigma_i$, such that $p_m \geq p_i, \forall i$. Let $I = \{1, 2, 3, 4\} \setminus \{m\}$ denote the set of the three indices of the probabilities that are not the largest. Then, the channel is antidegradable if and only if

$$p_i + p_j + p_k + \sqrt{p_i p_j} + \sqrt{p_j p_k} + \sqrt{p_k p_i} \geq \frac{1}{2}, \quad (\text{E5})$$

with $i, j, k \in I$ and $i \neq j \neq k$.

A visualization of the set of all antidegradable and EB channels is shown in Fig. 7. We see that the set of entanglement-breaking channels is a strict subset of the antidegradable channels. To quantify the size of these sets, we compare their volumes. The volume of the tetrahedron of all Pauli channels is $V_P = \frac{1}{6}$. The volume of the entanglement-breaking channels is found to be $V_{EB} = 0.083$, while the volume of the set of antidegradable channels is $V_{AD} = 0.151$. As ratios of the volume of all Pauli channels, 50% are entanglement breaking, while 90.7% are antidegradable.

Thus, 90.7% of unital qubit channels cannot be distilled with any protocol with definite causal order but can be distilled with our proposed protocols with SWITCH_n , with $n \geq 2$. At the same time, using the quantum switch, only the intersection of the zero-capacity set with the boundary can be distilled, a set of measure zero in the three-dimensional set of Pauli channels.

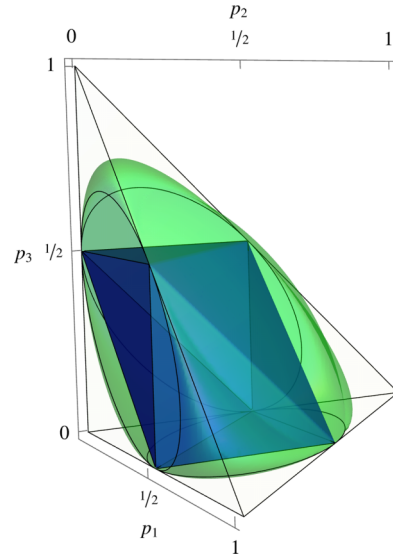


FIG. 7. The set of antidegradable channels is shown in green. The set of the entanglement-breaking channels is shown in blue.

APPENDIX F: DISTILLATION RATES FOR n -ORDER SUPERSWITCHES

As discussed in Appendix B, for any Pauli channel $\mathcal{E}_{\vec{p}}$, four copies of which are combined in the second-order superswitch, there is a nonzero probability $q_{- - +}^{(2)} > 0$ that the outcome “ $- - +$ ” is obtained. By definition, the distillation rate $\mathcal{R}^{(2)}$ is the probability that the outcome that leads to the identity channel is obtained, i.e., “ $- - +$,” and thus for the second-order superswitch is

$$\mathcal{R}^{(2)} = q_{- - +} = 4(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2). \quad (\text{F1})$$

The observation regarding the second-order superswitch and the ability to lead to the identity channel with nonzero probability for any channel holds for higher-order superswitches as well. Specifically, at order n there are $2^n - 1$ ancilla qubits. The last ancilla controls the order of the inside two $(n - 1)$ -order superswitches, while the second and third from the end control the ordering of lower-order superswitches inside these two $(n - 1)$ -order superswitches. It follows that whatever the previous outcomes $o_1, o_2, \dots$ on the rest of the ancilla qubits, the resulting channel is the identity channel, with some probability $0 \leq q_{o_1, \dots, - - +}^{(n)} \leq 1$. We note that some sequences of outcomes may have probability of occurrence 0. This argument ignores the possibility of unitary channels that can arise on the face of the tetrahedron and thus applies only in the interior. The complete derivation of the distillation rates that include the faces is given in Appendix J. With this in mind, the distillation rate $\mathcal{R}^{(n)}$ of an n -order superswitch is then

$$\mathcal{R}^{(n)} = \sum_{o_1, \dots, \pm} \Pr o_1, \dots, o_{2^n-4}, -, -, +. \quad (\text{F2})$$

For the depolarization channel, the distillation rates of up to order $n = 5$ are

$$\begin{aligned}\mathcal{R}^{(2)} &= \frac{3p^4}{64}, \quad \mathcal{R}^{(3)} = \frac{3}{64}(-2+p)p^4 + \frac{9p^8}{4096}, \\ \mathcal{R}^{(4)} &= 12p^4 - 72p^5 + 210p^6 - 390p^7 + \frac{4095p^8}{8} - \frac{999p^9}{2} + \frac{11877p^{10}}{32} - \frac{6759p^{11}}{32} + \frac{746673p^{12}}{8192} - \frac{118833p^{13}}{4096} \\ &\quad + \frac{105651p^{14}}{16384} - \frac{14673p^{15}}{16384} + \frac{981201p^{16}}{1677216},\end{aligned}\quad (\text{F3})$$

and

$$\begin{aligned}\mathcal{R}^{(5)} &= 192p^4 - 2688p^5 + 19488p^6 - 96096p^7 + 359052p^8 - 1076304p^9 + 2679264p^{10} - 5663592p^{11} \\ &\quad + \frac{660468723p^{12}}{64} - \frac{524016063p^{13}}{32} + \frac{2916648909p^{14}}{128} - \frac{3575251239p^{15}}{128} + \frac{247771513233p^{16}}{8192} \\ &\quad - \frac{29664177201p^{17}}{1024} + \frac{40211190275p^{18}}{16384} - \frac{301008655485p^{19}}{16384} + \frac{25426529527977p^{20}}{2097152} - \frac{7371155519931p^{21}}{1048576} \\ &\quad + \frac{14948591861835p^{22}}{4194304} - \frac{6585748985817p^{23}}{4194304} + \frac{40009926738525p^{24}}{67108864} - \frac{3238417329813p^{25}}{16777216} \\ &\quad + \frac{14102865567135p^{26}}{268435456} - \frac{3165750102405p^{27}}{268435456} + \frac{146071342742253p^{28}}{68719476736} - \frac{10160062539117p^{29}}{34359738368} \\ &\quad + \frac{4095741366951p^{30}}{137438953472} - \frac{266117803725p^{31}}{137438953472} + \frac{17141597128353p^{32}}{281474976710656}.\end{aligned}\quad (\text{F4})$$

APPENDIX G: THE CHOI-JAMIOŁKOWSKI ISOMORPHISM AND ENTANGLEMENT SHARING

Our results can be directly translated to an entanglement sharing scenario. In this setting, a party prepares some entangled state, keeps one part of that state, and sends the other to the second party through some channel $\mathcal{E}$. Clearly, if the channel is entanglement breaking, this task cannot be performed. In addition, unless the channel is the identity, $\mathbf{Id}$, or a unitary channel, sharing a maximally entangled state is also impossible. Our main results, however, show that this is possible given access to at least the second-order superswitch. This follows from the Choi-Jamiołkowski isomorphism [39,40,53], which establishes a correspondence between bipartite states and channels. Specifically, a map between channels and bipartite states is established through the equation

$$\rho = (\mathbf{Id} \otimes \mathcal{E})[|\Psi\rangle\langle\Psi|], \quad (\text{G1})$$

where $|\Psi\rangle\langle\Psi|$ is a maximally entangled state. Pauli channels are mapped to Bell-diagonal states, which are mixtures of the four entangled states. Specifically, a Pauli channel $\mathcal{P}$ is mapped to the Bell-diagonal state

$$\begin{aligned}\rho_B &= p_0|\Phi^+\rangle\langle\Phi^+| + p_1|\Psi^+\rangle\langle\Psi^+| \\ &\quad + p_2|\Psi^-\rangle\langle\Psi^-| + p_3|\Phi^-\rangle\langle\Phi^-|,\end{aligned}\quad (\text{G2})$$

where

$$\begin{aligned}|\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad |\Phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)\end{aligned}\quad (\text{G3})$$

are the four Bell states. We note that a Bell-diagonal state is separable if

$$p_{\max} = \max_i \{p_i\} \leq \frac{1}{2}. \quad (\text{G4})$$

The depolarization channel, defined with parameters $p_1 = p_2 = p_3 = \frac{p}{4}$, is mapped to

$$\rho = (1-p)|\Phi^+\rangle\langle\Phi^+| + p\frac{\mathbb{1}_4}{4}, \quad (\text{G5})$$

which is a Werner state [54]

$$\rho_W = \lambda|\Psi^-\rangle\langle\Psi^-| + (1-\lambda)\frac{\mathbb{1}_4}{4}, \quad (\text{G6})$$

up to a unitary transformation. It is known that Werner states are entangled whenever $\lambda > \frac{1}{3}$.

Through the Choi-Jamiołkowski isomorphism, we show that results on channels are translated to results on bipartite states through the quantum switch, which gives

$$\begin{aligned}\tilde{S}_\omega(\mathcal{F}, \mathcal{F}) &= \frac{1}{4} \sum_{i,j} (\{F_i, F_j\} \rho \{F_i, F_j\}^\dagger \otimes \omega \\ &\quad + [F_i, F_j] \rho [F_i, F_j]^\dagger \otimes Z\omega Z),\end{aligned}\quad (\text{G7})$$

where the channel $\mathcal{F}$ is acting on bipartite states ρ_{AB} and is given by $\mathcal{F} = \mathbb{1} \otimes \mathcal{E}$, where $\mathcal{E}$ is a channel acting on system B . Then, the Kraus operators are explicitly $F_i = \mathbb{1} \otimes E_i$. If $\mathcal{E}$ is a Pauli channel, we find that the result of applying the quantum switch is

$$\tilde{S}_\omega(\mathcal{E}, \mathcal{E}) = q_+ \rho_{AB}^{(+)} \otimes \omega_+ + q_- \rho_{AB}^{(-)} \otimes \omega_-, \quad (\text{G8})$$

where the bipartite states $\rho_{AB}^{(+)}$ and $\rho_{AB}^{(-)}$ are defined as

$$\begin{aligned}\rho_{AB}^{(+)} &= (p_0^2 + p_1^2 + p_2^2 + p_3^2)|\Phi^+\rangle\langle\Phi^+| + 2p_0(p_1|\Psi^+\rangle\langle\Psi^+| \\ &\quad + p_2|\Psi^-\rangle\langle\Psi^-| + p_3|\Phi^-\rangle\langle\Phi^-|)/q_+, \\ \rho_{AB}^{(-)} &= (2p_1p_2|\Phi^-\rangle\langle\Phi^-| + 2p_2p_3|\Psi^+\rangle\langle\Psi^+| \\ &\quad + 2p_3p_1|\Psi^-\rangle\langle\Psi^-|)/q_-, \end{aligned} \quad (\text{G9})$$

with

$$q_- = 2(p_1p_2 + p_2p_3 + p_3p_1), \quad q_+ = 1 - q_-, \quad (\text{G10})$$

and $\omega_+ = \omega$ and $\omega_- = Z\omega Z$. The choice $\omega = |+\rangle\langle+|$ leads to $\omega_{\pm} = |\pm\rangle\langle\pm|$, and thus a measurement in the $|\pm\rangle$ basis can separate the two states completely. A similar picture holds for the higher-order superswitches. For instance, the second-order superswitch follows directly by applying Eq. (G2) to the expressions [Eqs. (B13) and (B14)].

APPENDIX H: TWIRLING CHANNELS AND APPLICATION TO SUPERSWITCHES

Given a channel $\mathcal{E}$, twirling is a sort of averaging with random applications of unitaries over the Haar measure [43,44]. It can be seen as a supermap that maps any channel to the depolarizing channel. Specifically, it is defined through the equation

$$\begin{aligned}\mathcal{T}[\mathcal{E}](\rho) &= \int dU U^\dagger \mathcal{E}(U\rho U^\dagger)U \\ &= (1 - \eta_{\mathcal{E}})\rho + \eta_{\mathcal{E}} \frac{\mathbb{1}}{2} \equiv \mathcal{D}_{\eta_{\mathcal{E}}}(\rho) \end{aligned} \quad (\text{H1})$$

in the case with $d = 2$. If the channel is expressed with Kraus operators as $\mathcal{E}(\rho) = \sum K_i \rho K_i^\dagger$, then the depolarizing parameter is obtained as

$$1 - \eta_{\mathcal{E}} = \frac{\sum_i |\text{tr} K_i^2| - 1}{3}. \quad (\text{H2})$$

Note that in the special case of a Pauli channel $\mathcal{P}$, the last equation reduces to

$$1 - \eta_{\mathcal{P}} = \frac{4p_0 - 1}{3}. \quad (\text{H3})$$

A *unitary 2-design* [44] is a finite set of unitary matrices $U\mathcal{D} = \{U_i\}_{i=1,\dots,N}$ that can perform twirling. That is,

$$\frac{1}{N} \sum_i U_i^\dagger \mathcal{E}(U_i \rho U_i^\dagger) U_i \equiv \mathcal{T}_{\mathcal{E}}(\rho) = (1 - \eta_{\mathcal{E}})\rho + \eta_{\mathcal{E}} \frac{\mathbb{1}}{2}. \quad (\text{H4})$$

In dimension $d = 2$, the Clifford group is an example of a unitary 2-design. It consists of 24 elements, but there exist unitary 2-designs with fewer elements, with 10 being the lower bound. Moreover, it is known that for the restricted class of Pauli channels, three unitaries are sufficient to perform twirling, making this protocol practical [55]. The above machinery is important for our purposes as it allows us to turn any channel into a depolarization one and thus the statements made for Pauli channels apply implicitly to nonunitary channels as well. For a single use of a channel, and a unitary-2 design with N elements, in practice one can perform twirling by

- (1) producing N copies of the state,
- (2) $\forall i$ applying a different U_i to each copy,

- (3) sending each $U_i \rho U_i^\dagger$ through a copy of the channel $\mathcal{E}$,
- (4) applying a $U_i^\dagger$ to each copy with the same index,
- (5) using each copy for whatever purpose (e.g., measuring), and
- (6) averaging over outcomes over all copies, thus obtaining $\mathcal{T}[\mathcal{E}](\rho) = \mathcal{D}_{\eta}(\rho)$.

However, as the superswitches combine multiple copies of the channel, this procedure needs to be adapted. We describe the modified implementation for the case of the quantum switch, which combines two copies of the channel, and the generalization to higher-order superswitches is straightforward. Let us denote by $\mathcal{E}_i$ the channels $\mathcal{E}_i = \mathcal{U}_i^\dagger \circ \mathcal{E} \circ \mathcal{U}_i$, where $\mathcal{U}_i(\rho) = U_i \rho U_i^\dagger$ denotes a unitary channel and $\mathcal{U}_i^\dagger$ its adjoint. Each channel $\mathcal{E}_i$ has Kraus operators $E_{ik} = U_i^\dagger E_k U_i$, where E_k are the Kraus operators of $\mathcal{E}$. Then, it follows that

$$\begin{aligned}\frac{1}{N^2} \sum_{i,j} \mathcal{E}_i \circ \mathcal{E}_j(\rho) &= \frac{1}{N} \sum_i \mathcal{E}_i \left(\frac{1}{N} \sum_j \mathcal{E}_j(\rho) \right) \\ &= \frac{1}{N} \sum_i \mathcal{E}_i \left(\frac{1}{N} \sum_j U_j^\dagger \mathcal{E}(U_j \rho U_j^\dagger) U_j \right) \\ &= \frac{1}{N} \sum_i U_i^\dagger \mathcal{E}(U_i \mathcal{D}_{\eta_{\mathcal{E}}}(\rho) U_i^\dagger) U_i \\ &= \mathcal{D}_{\eta_{\mathcal{E}}}(\mathcal{D}_{\eta_{\mathcal{E}}}(\rho)), \end{aligned} \quad (\text{H5})$$

where we have used the linearity of channels. What this means is that by making all possible combinations of $\mathcal{E}_i$ and $\mathcal{E}_j$ on different copies of the state and the channels, and averaging over them, we can create the composition of the two twirled channels. As for the way to compute the switch of the twirled channels in practice, this can be done by

- (1) producing N^2 copies of the state,
- (2) for each value of i, j combining the channels $\mathcal{E}_i$ and $\mathcal{E}_j$ in a quantum switch,
- (3) sending each copy through a switch $\mathcal{S}(\mathcal{E}_i, \mathcal{E}_j)$,
- (4) using each copy for whatever purpose (e.g., measuring), and
- (5) averaging outcomes over all N^2 copies.

Note, however, that there is the following complication. As a measurement is made on the ancilla for each of the N^2 copies, same measurement outcomes should be grouped and averaged over. That means that for each value of i, j at least one $+$ outcome needs to be obtained before the averaging to produce the C_+ channel, and similarly for $-$. Thus, given the randomness of the outcomes after measurements in the ancilla, more repetitions are necessary. For instance, to replicate the C_+ channel one would need to keep repeating the process for each i, j until the $+$ outcome is obtained, which clearly has a high resource cost.

However, irrespective of its inefficiency, this effectively evaluates the quantum switch of the two depolarization channels that follow from twirling, i.e., $\mathcal{S}(\mathcal{D}_{\eta_{\mathcal{E}}}, \mathcal{D}_{\eta_{\mathcal{E}}})(\rho)$. Thus, results on the depolarization channel can be effectively applied to any channel, however, at the significant resource cost of multiple copies of the state ρ , the channel $\mathcal{E}$, and the need for one switch for each copy of the state, specifically N^2 copies of the state, $2N^2$ copies of the channel, and N^2 quantum switches. For example, for a general unknown

channel in $d = 2$, assuming the existence of a minimal unitary 2-design with 10 elements, one would need to implement 100 switches in parallel. In practice, however, this number can be significantly smaller if the class of channels is restricted. That is, channel-specific twirling in some cases exists that can be used to significantly reduce this overhead. For instance, a Pauli channel can be twirled with only three unitary matrices [55], significantly reducing the cost from 10. In this case, only 9 copies of the state, 18 copies of the channel, and 9 quantum switches are sufficient.

The generalization to the case of a superswitch of order n is straightforward. The n -order superswitch combines 2^n copies of the channel and one can twirl and thus depolarize each of them by

- (1) producing N^{2^n} copies of the state,
- (2) for each value of the 2^n indices $i_1, \dots, i_{2^n}$ combining the channels $\mathcal{E}_{i_1}, \dots, \mathcal{E}_{i_{2^n}}$ in an n -order superswitch,
- (3) sending each copy through the superswitch $\mathcal{S}^{(n)}(\mathcal{E}_{i_1}, \dots, \mathcal{E}_{i_{2^n}})$,
- (4) using each copy for whatever purpose (e.g., measuring), and
- (5) averaging outcomes over all N^{2^n} copies.

This effectively evaluates the n -order superswitch of the 2^n depolarization channels that follow from twirling, i.e., $\mathcal{S}^{(n)}(\mathcal{D}_{\eta_{\mathcal{E}}}, \dots, \mathcal{D}_{\eta_{\mathcal{E}}})(\rho)$. Clearly, this becomes impractical very quickly as N^{2^n} copies of the state, $2^n N^{2^n}$ copies of the channel, and N^{2^n} n -order superswitches are necessary to implement this. However, once again this overhead may be reduced by channel-specific twirling.

Finally, in view of the main results, to distil the channel, it is only necessary to implement the second-order superswitch and obtain the sequence of outcomes “− − +.” Thus, for channel distillation, a minimal protocol is given in the following steps:

- (1) Produce a copy of the state ρ to be sent over the channel.
- (2) For each value of the four indices i, j, k , and l , combine the channels $\mathcal{E}_i, \mathcal{E}_j, \mathcal{E}_k$, and $\mathcal{E}_l$ in a first-order superswitch and measure the three ancillas.
- (3) If the obtained outcome is different from “− − +,” repeat. Otherwise, use the obtained state for whatever purpose (e.g., measuring) and proceed to a new combination of values of i, j, k , and l .
- (4) When all combinations of i, j, k , and l have been exhausted and an outcome “− − +” has been obtained for each one of them, average the outcomes on the states over all copies with different indices, thus obtaining the original state ρ .

APPENDIX I: RESULTS 1 AND 2 DO NOT HOLD IN $d > 2$

In this Appendix, we show that our results do not hold for all unital channels in any dimension $d = 2^n$ with $n > 1$. We recall that the reason that this was possible in $d = 2$ was based on two facts: (1) for any channel, a minus result on the ancillas gives a channel with $p_0 = 0$, and (2) for any Pauli channel with $p_0 = 0$, a plus outcome leads to the identity channel. We will show that while (1) still holds, (2) fails.

The general expression for the superswitches [Eq. (B2)] involves terms with nested commutators and anticommutators of Kraus operators of the channels. A recursive procedure

to evaluate higher-order superswitches from lower-order ones was given in Ref. [21]. Effectively, commutators and anticommutators of Kraus operators of a Pauli channel map to another Pauli channel, as in the expression for the quantum switch [Eq. (G7)].

In order to evaluate the resulting channels, let $\mathcal{E}(\rho) = \sum_i \tilde{p}_{1,i} \sigma_i \rho \sigma_i$ and $\mathcal{F}(\rho) = \sum_i \tilde{p}_{2,i} \sigma_i \rho \sigma_i$ denote two Pauli channels with probability vectors $\tilde{p}_i = \{\alpha_i, \beta_i, \gamma_i, \delta_i\}$, where $\vec{\sigma} = \{\mathbb{1}, X, Y, Z\}$ is the vector of Pauli matrices. We gather all anticommutators and commutators of Kraus operators in two matrices with operator-valued elements, $\mathcal{A}[\mathcal{E}, \mathcal{F}]$ and $\mathcal{C}[\mathcal{E}, \mathcal{F}]$, respectively, given by

$$\mathcal{A}[\mathcal{E}, \mathcal{F}] = 2 \begin{pmatrix} \sqrt{\alpha_1 \alpha_2} \mathbb{1} & \sqrt{\alpha_1 \beta_2} X & \sqrt{\alpha_1 \gamma_2} Y & \sqrt{\alpha_1 \delta_2} Z \\ \sqrt{\beta_1 \alpha_2} X & \sqrt{\beta_1 \beta_2} \mathbb{1} & 0 & 0 \\ \sqrt{\gamma_1 \alpha_2} Y & 0 & \sqrt{\gamma_1 \gamma_2} \mathbb{1} & 0 \\ \sqrt{\delta_1 \alpha_2} Z & 0 & 0 & \sqrt{\delta_1 \delta_2} \mathbb{1} \end{pmatrix} \quad (11)$$

and

$$\mathcal{C}[\mathcal{E}, \mathcal{F}] = 2i \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\beta_1 \gamma_2} Z & -\sqrt{\beta_1 \delta_2} Y \\ 0 & -\sqrt{\gamma_1 \beta_2} Z & 0 & \sqrt{\gamma_1 \delta_2} X \\ 0 & \sqrt{\delta_1 \beta_2} Y & -\sqrt{\delta_1 \gamma_2} X & 0 \end{pmatrix}. \quad (12)$$

Then, the two terms in the definition of the quantum switch [Eq. (B3)] are equivalent to

$$\begin{aligned} \mathfrak{a}[\mathcal{E}, \mathcal{F}] &\equiv \frac{1}{4} \sum_{i,j} \{E_i, F_j\} \rho \{E_i, F_j\}^\dagger \\ &= e^\top \cdot (\mathcal{A}[\mathcal{E}, \mathcal{F}] \circ \rho \circ \mathcal{A}[\mathcal{E}, \mathcal{F}]^*) \cdot e, \\ \mathfrak{c}[\mathcal{E}, \mathcal{F}] &\equiv \frac{1}{4} \sum_{i,j} [E_i, F_j] \rho [E_i, F_j]^\dagger \\ &= e^\top \cdot (\mathcal{C}[\mathcal{E}, \mathcal{F}] \circ \rho \circ \mathcal{C}[\mathcal{E}, \mathcal{F}]^*) \cdot e, \end{aligned} \quad (13)$$

where $e = (1, 1, 1, 1)$ denotes a vector of 1s and “ $\cdot$ ” denotes usual matrix multiplication. We have introduced two operations: (1) “ $\circ$ ” that takes two 4×4 operator-valued matrices on the left and right sides and a density matrix in the middle and returns a 4×4 operator-valued matrix, and (2) “ $*$ ” that denotes taking the conjugate transpose of each element of the operator-valued matrix. In detail, the first operation consists of elementwise multiplication of the two outside matrices while “sandwiching” the density matrix ρ with the operators on the left and right, keeping the order. For example, the third element of the first row of $\mathcal{A} \circ \rho \circ \mathcal{A}^*$ would give the contribution $[\mathcal{A} \circ \rho \circ \mathcal{A}^*]_{13} = \alpha_1 \gamma_2 Y \rho Y$. The other operation, $*$, is unimportant in the case we are considering, as conjugate transposes of Pauli matrices leave them unchanged. By taking left and right matrix products with the vector with unit entries e , we effectively get all terms in the summation.

With the introduction of this concise notation, we readily find

$$\begin{aligned} \mathfrak{a}[\mathcal{E}, \mathcal{F}] &= (\alpha_1\alpha_2 + \beta_1\beta_2 + \gamma_1\gamma_2 + \delta_1\delta_2)\rho \\ &\quad + (\alpha_1\beta_2 + \beta_1\alpha_2)X\rho X + (\alpha_1\gamma_2 + \gamma_1\alpha_2)Y\rho Y \\ &\quad + (\alpha_1\delta_2 + \delta_1\alpha_2)Z\rho Z, \\ \mathfrak{c}[\mathcal{E}, \mathcal{F}] &= (\gamma_1\delta_2 + \delta_1\gamma_2)X\rho X + (\delta_1\beta_2 + \beta_1\delta_2)Y\rho Y \\ &\quad + (\beta_1\gamma_2 + \gamma_1\beta_2)Z\rho Z. \end{aligned} \quad (14)$$

Note that the expressions do not give channels but channels multiplied by probabilities.

The generalization to higher-order superswitches is straightforward. For instance, the channel of interest corresponding to the outcomes “− +” in the case of the second-order superswitch is

$$\begin{aligned} &\frac{1}{64} \sum_{i,j,k,l} \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\} \rho \{[E_i, F_j], [\tilde{E}_k, \tilde{F}_l]\}^\dagger \\ &= e^\top \cdot (\mathcal{A}[\mathfrak{c}[\mathcal{E}, \mathcal{F}], \mathfrak{c}[\tilde{\mathcal{E}}, \tilde{\mathcal{F}}]] \circ \rho \circ \mathcal{A}[\mathfrak{c}[\mathcal{E}, \mathcal{F}], \mathfrak{c}[\tilde{\mathcal{E}}, \tilde{\mathcal{F}}]]^*) \\ &\quad \cdot e. \end{aligned} \quad (15)$$

Letting $\mathcal{E} = \mathcal{F} = \tilde{\mathcal{E}} = \tilde{\mathcal{F}} = \sum_i p_i \sigma_i \rho \sigma_i$, we readily obtain

$$C_{--+}^{(2)} = 4(p_1^2 p_2^2 + p_2^2 p_3^2 + p_3^2 p_1^2) \rho = q_{--+}^{(2)} \mathbf{Id}. \quad (16)$$

A crucial feature for this is the fact that for two channels with $\alpha_1 = \alpha_2 = 0$, the only terms remaining in the matrix of anticommutators $\mathcal{A}[\mathcal{E}, \mathcal{F}]$ in Eq. (11) involve the identity matrix. However, when moving to Pauli channels in dimensions 2^l with $l > 1$, all Kraus operators involve operators of the form $\sigma_i \otimes \sigma_j \otimes \dots$. This implies that the matrix of anticommutators $\mathfrak{a}[\mathcal{E}, \mathcal{F}]$ does not have the simple form of Eq. (11). Specifically, even when the portion of the identity of the two channels $\mathcal{E}$ and $\mathcal{F}$ is equal to zero, there will remain terms that involve operators that are not proportional to $\mathbb{1} \otimes \mathbb{1} \dots \otimes \mathbb{1}$. For instance, in $d = 4$ we have the anticommutator

$$\{\mathbb{1} \otimes X, X \otimes \mathbb{1}\} = 2X \otimes X \quad (17)$$

as well as similar terms with X replaced with Y or Z , among others. Thus, it is clear that the crucial feature that enabled distillation to the identity channel in $d = 2$ fails to hold in $d = 4$. Moreover, such terms are also included trivially in any dimension $d = 2^n$ with any $n > 2$ as, for example,

$$\begin{aligned} &\{\mathbb{1} \otimes \dots \otimes \mathbb{1} \otimes X, \mathbb{1} \otimes \dots \otimes X \otimes \mathbb{1}\} \\ &= \mathbb{1} \otimes \dots \otimes \{\mathbb{1} \otimes X, X \otimes \mathbb{1}\} \\ &= 2(\mathbb{1} \otimes \dots \otimes X \otimes X). \end{aligned} \quad (18)$$

It follows that in any dimension $d \neq 2$, Results 1 and 2 do not hold for all Pauli channels.

APPENDIX J: RECURRENCE RELATIONS AND ASYMPTOTIC DISTILLATION RATE

1. Channels in the interior of the tetrahedron

In this Appendix, we show that the distillation rate after the action of a superswitch of order n on 2^n copies of a Pauli channel $\mathcal{P}_{\vec{p}}$ can be evaluated given knowledge of the action of the superswitch at order $n - 1$. Thus, we construct recurrence relations that can be iterated to find the action of the

superswitch at any order. In addition, by analyzing the fixed points of the recurrence relations, we can find the asymptotic behavior that follows from the superswitches and thus obtain the asymptotic distillation rate for certain channels.

We use the following definitions. We denote a Pauli channel with $\mathcal{P}_{\vec{p}}(\rho) = p_0\rho + p_1X\rho X + p_2Y\rho Y + p_3Z\rho Z$ and for notational simplicity we may represent it by its probability vector $\vec{p} = \{p_0, p_1, p_2, p_3\}$. We denote with $\mathcal{N}_{\vec{q}}$ a channel that has zero portion of the identity, that is, the channel $\mathcal{N}_{\vec{q}}(\rho) = q_1X\rho X + q_2Y\rho Y + q_3Z\rho Z$. We recall Eq. (14) of Appendix I, and rewrite concisely in terms of the two vectors $\vec{p}$ and $\vec{q}$ of two Pauli channels $\mathcal{E}_{\vec{p}}$ and $\mathcal{F}_{\vec{q}}$ as

$$\begin{aligned} \mathfrak{a}[\mathcal{E}, \mathcal{F}] &= \Pr(\mathfrak{a}) \left\{ \frac{\sum_{i=0}^3 p_i q_i}{\Pr(\mathfrak{a})}, \frac{p_0 q_1 + p_1 q_0}{\Pr(\mathfrak{a})}, \right. \\ &\quad \times \left. \frac{p_0 q_2 + p_2 q_0}{\Pr(\mathfrak{a})}, \frac{p_0 q_3 + p_3 q_0}{\Pr(\mathfrak{a})} \right\}, \\ \mathfrak{c}[\mathcal{E}, \mathcal{F}] &= \Pr(\mathfrak{c}) \left\{ 0, \frac{p_1 q_2 + p_2 q_1}{\Pr(\mathfrak{c})}, \frac{p_2 q_3 + p_3 q_2}{\Pr(\mathfrak{c})}, \right. \\ &\quad \times \left. \frac{p_3 q_1 + p_1 q_3}{\Pr(\mathfrak{c})} \right\}, \end{aligned} \quad (J1)$$

where

$$\begin{aligned} \Pr(\mathfrak{c}) &= p_1 q_2 + p_2 q_1 + p_2 q_3 + p_3 q_2 + p_3 q_1 + p_1 q_3, \\ \Pr(\mathfrak{a}) &= 1 - \Pr(\mathfrak{c}). \end{aligned} \quad (J2)$$

In turn, we can rewrite this as

$$\mathfrak{a}[\mathcal{E}, \mathcal{F}] = \Pr(\mathfrak{a}) \mathcal{P}_{\vec{f}(\vec{p}, \vec{q})}, \quad \mathfrak{c}[\mathcal{E}, \mathcal{F}] = \Pr(\mathfrak{c}) \mathcal{N}_{\vec{g}(\vec{p}, \vec{q})}, \quad (J3)$$

with

$$\begin{aligned} \vec{f}(\vec{p}, \vec{q}) &= \frac{1}{\Pr(\mathfrak{a})} \left\{ \sum_{i=0}^3 p_i q_i, p_0 q_1 + p_1 q_0, \right. \\ &\quad \times \left. p_0 q_2 + p_2 q_0, p_0 q_3 + p_3 q_0 \right\}, \\ \vec{g}(\vec{p}, \vec{q}) &= \frac{1}{\Pr(\mathfrak{c})} \{0, p_1 q_2 + p_2 q_1, p_2 q_3 + p_3 q_2, \\ &\quad \times p_3 q_1 + p_1 q_3\}. \end{aligned} \quad (J4)$$

We see that when two Pauli channels are combined in a quantum switch, the anticommutator term that corresponds to the “+” outcome after a measurement on the ancilla leads to an updated Pauli channel with probability vector $\vec{f}(\vec{p}, \vec{q})$, while the commutator term that corresponds to the “−” outcome after a measurement on the ancilla leads to a Pauli channel with probability vector $\vec{g}(\vec{p}, \vec{q})$, which has zero portion of the identity. In addition, the identity channel can only be obtained if $p_0 = q_0 = 0$. This means that at each order only outcomes from the previous order that lead to channels with zero portion of the identity can lead to the identity at the current order. As a result, at order n we need to track and group together three terms that lead to different behaviors:

- (1) Outcomes that lead to the identity channel $\mathbf{Id}$.

(2) Outcomes that lead to channels with zero portion of the identity, $\mathcal{N}_{\vec{q}^{(n)}}$.

(3) Outcomes that lead to Pauli channels that do not belong to either of the two aforementioned classes, $\mathcal{P}_{\vec{p}^{(n)}}$.

The total resulting channel at order n of the superswitch can be written as

$$\mathcal{S}^{(n)} = \alpha^{(n)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n)} \mathcal{N}_{\vec{q}^{(n)}} \otimes \rho_{\mathcal{N}} + \gamma^{(n)} \mathcal{P}_{\vec{p}^{(n)}} \otimes \rho_{\mathcal{P}}. \quad (\text{J5})$$

By $\rho_{\mathbf{Id}}$, $\rho_{\mathcal{N}}$, and $\rho_{\mathcal{P}}$ we denote the unnormalized states with orthogonal supports that gather the outcomes $o_{\mathbf{Id}}$, $o_{\mathcal{N}}$, and $o_{\mathcal{P}}$ of measurements on the ancillas that lead to the identity channel $\mathbf{Id}$, a channel with zero portion of the identity $\mathcal{N}$, i.e., $q_0^{(n)} = 0$, and a general Pauli channel with $p_0^{(n)} \neq 0, 1$, respectively. For instance, in the case of the first-order superswitch, it follows from Eq. (B13) that

$$\begin{aligned} \rho_{\mathbf{Id}} &= | - - + \rangle \langle - - + |, \\ \rho_{\mathcal{N}} &= | + + - \rangle \langle + + - | + | - + - \rangle \langle - + - | \\ &\quad + | + - - \rangle \langle + - - | + | - - - \rangle \langle - - - |, \\ \rho_{\mathcal{P}} &= \mathbb{1} - \rho_{\mathbf{Id}} - \rho_{\mathcal{N}} = (| + + + \rangle \langle + + + | \\ &\quad + | - + + \rangle \langle - + + | + | + - + \rangle \langle + - + |). \end{aligned} \quad (\text{J6})$$

The fact that it is possible to summarize all different channels into essentially three channels follows from the fact that for two Pauli channels $\mathcal{P}_{\vec{p}}$ and $\mathcal{P}_{\vec{q}}$,

$$\sum_i \lambda_i \mathcal{P}_{\vec{p}_i} = \left(\sum_i \lambda_i \right) \mathcal{P}_{\vec{p}'}, \quad \vec{p}' = \frac{\sum_i \lambda_i \vec{p}_i}{\sum_i \lambda_i}. \quad (\text{J7})$$

Given that superswitches are switches of switches and that the action of the superswitch at order n is assumed to take the form in Eq. (J5), we can explicitly evaluate the superswitch at order $n + 1$. In practice, that means that for each combination of terms we need to evaluate the contribution from the anti-commutator and commutator contribution from Eq. (3). There are nine possible combinations, each of which leads to two possible outcomes, so in principle we will have 18 terms at order $n + 1$. However, by noticing that combining the identity with any channel leads only to one outcome, i.e., the channel itself, and that the ordering of combining two channels, i.e., $\alpha[\mathcal{E}, \mathcal{F}] = \alpha[\mathcal{F}, \mathcal{E}]$ and $\mathcal{C}[\mathcal{E}, \mathcal{F}] = \mathcal{C}[\mathcal{F}, \mathcal{E}]$ does not matter, we find nine unique terms. Explicitly,

$$\begin{aligned} \mathcal{S}^{(n+1)} &= \left((\alpha^{(n)})^2 + \frac{(\beta^{(n)})^2}{3} \right) \mathbf{Id} \otimes \rho_{\mathbf{Id}} \\ &\quad + \left(2\alpha^{(n)}\beta^{(n)}\mathcal{N}_{\vec{q}^{(n)}} + \frac{2(\beta^{(n)})^2}{3}\mathcal{N}_{\vec{q}_{\frac{4}{3}}} \right. \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{p}^{(n)}])\mathcal{N}_{\vec{g}(\vec{p}^{(n)}, \vec{p}^{(n)})} \\ &\quad + 2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{q}^{(n)}])\mathcal{N}_{\vec{g}(\vec{p}^{(n)}, \vec{q}^{(n)})} \left. \right) \otimes \rho_{\mathcal{N}} \\ &\quad + (2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{q}^{(n)}])\mathcal{P}_{\vec{f}(\vec{p}^{(n)}, \vec{q}^{(n)})} \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{p}^{(n)}])\mathcal{P}_{\vec{f}(\vec{p}^{(n)}, \vec{p}^{(n)})} \\ &\quad + 2\alpha^{(n)}\gamma^{(n)}\mathcal{P}_{\vec{p}^{(n)}} \otimes \rho_{\mathcal{P}}, \end{aligned} \quad (\text{J8})$$

where $\vec{q}_{\frac{4}{3}} = \{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$. The last expression is of the form

$$\mathcal{S}^{(n+1)} = \alpha^{(n+1)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n+1)} \mathcal{N}_{\vec{q}^{(n+1)}} \otimes \rho_{\mathcal{N}} + \gamma^{(n+1)} \mathcal{P}_{\vec{p}^{(n+1)}} \otimes \rho_{\mathcal{P}}, \quad (\text{J9})$$

where

$$\begin{aligned} \alpha^{(n+1)} &= (\alpha^{(n)})^2 + \frac{(\beta^{(n)})^2}{3}, \\ \beta^{(n+1)} &= 2\alpha^{(n)}\beta^{(n)} + \frac{2(\beta^{(n)})^2}{3} \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{p}^{(n)}]) \\ &\quad + 2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{q}^{(n)}]), \\ \gamma^{(n+1)} &= 2\alpha^{(n)}\gamma^{(n)} + 2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{q}^{(n)}]) \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{p}^{(n)}]), \\ \vec{q}^{(n+1)} &= \left(2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{q}^{(n)}])\vec{g}(\vec{p}^{(n)}, \vec{q}^{(n)}) \right. \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{C}[\vec{p}^{(n)}, \vec{p}^{(n)}])\vec{g}(\vec{p}^{(n)}, \vec{p}^{(n)}) \\ &\quad + 2\alpha^{(n)}\beta^{(n)}\vec{q}^{(n)} + \left. \frac{2(\beta^{(n)})^2}{3}\vec{q}_{\frac{4}{3}} \right) / \beta^{(n+1)}, \\ \vec{p}^{(n+1)} &= (2\beta^{(n)}\gamma^{(n)} \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{q}^{(n)}])\vec{f}(\vec{p}^{(n)}, \vec{q}^{(n)}) \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathcal{A}[\vec{p}^{(n)}, \vec{p}^{(n)}])\vec{f}(\vec{p}^{(n)}, \vec{p}^{(n)}) \\ &\quad + 2\alpha^{(n)}\gamma^{(n)}\vec{p}^{(n)}) / \gamma^{(n+1)}. \end{aligned} \quad (\text{J10})$$

These equations effectively define recurrence relations that can be iterated to find the action of a superswitch of any order. The initial conditions for the recurrence relations are $\alpha^{(0)} = \beta^{(0)} = 0$, $\gamma^{(0)} = 1$, $\vec{q}^{(0)} = 0$, and $\vec{p}^{(0)} = \vec{p}$. Note that since the zeroth-order superswitch is the channel $\mathcal{P}_{\vec{p}}$ only without any superswitch acting, the index n starts from the value 0. We highlight that the parameter $\alpha^{(n)}$ gives the distillation rate at order n , which is zero in general for $n = 1$ (aside from channels with $p_0 = 0$). Evaluating the second-order superswitch by iterating, we find $\mathcal{R}^{(2)} = \alpha^{(2)} = 4(p_1 p_2 + p_2 p_3 + p_3 p_1)$. Higher orders can be evaluated iteratively, but we do not list the expressions as they become cumbersome.

Instead, we focus on a particular example, the depolarization channel. In this case, it can be seen that the parameters $\vec{q}^{(n)}$ remain constant for any value of n and take the value $\vec{q}^{(n)} = \vec{q}_{\frac{4}{3}} = \{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$. In addition, there is only one recurrence relation for $\vec{p}^{(n)}$ as the Pauli vector is that of a depolarization channel. Thus, the recurrence relations are significantly simplified. The distillation rates, $\mathcal{R}^{(n)} = \alpha^{(n)}$, can be obtained in a straightforward manner and the expressions for orders of up to 5 are given in Eqs. (F3) and (F4). Orders of up to 8 are plotted in Fig. 3.

Interestingly, in this case we can also analyze the asymptotic behavior by looking for the fixed points of the recurrence

relations. These follow by setting

$$\begin{aligned}\alpha^{(n+1)} &= \alpha^{(n)} := \alpha, \quad \beta^{(n+1)} = \beta^{(n)} := \beta, \\ \gamma^{(n+1)} &= \gamma^{(n)} := \gamma, \\ \bar{p}^{(n+1)} &= \bar{p}^{(n)} := \left\{1 - \frac{3p}{4}, \frac{p}{4}, \frac{p}{4}, \frac{p}{4}\right\},\end{aligned}\quad (\text{J11})$$

and explicitly we obtain

$$\begin{aligned}\alpha &= \alpha^2 + \frac{\beta^2}{3}, \\ \beta &= 2\alpha\beta + \frac{2\beta^2}{3} + 4\beta\gamma p + 6\gamma^2 p^2, \\ \gamma &= 2\alpha\gamma + 2\beta\gamma(1 - 2p) + \gamma^2(1 - 6p^2), \\ p &= \left(2\alpha\gamma p + 2\beta\gamma\left(\frac{1}{3} - p\right) + 2\gamma^2(1 - 3p)p\right) / \gamma.\end{aligned}\quad (\text{J12})$$

All solutions to these equations can be analytically obtained. Some correspond to the trivial solution of starting with the identity channel, $p = 0$. Another solution corresponds to the case of starting with the depolarization channel with $p = \frac{4}{3}$,

$$\alpha_{\frac{4}{3}}^{(\infty)} = \frac{1}{4}, \quad \beta_{\frac{4}{3}}^{(\infty)} = \frac{3}{4}, \quad \gamma_{\frac{4}{3}}^{(\infty)} = 0, \quad p_{\frac{4}{3}}^{(\infty)} = \frac{3 - \sqrt{3}}{6}. \quad (\text{J13})$$

The value of $\alpha_{\frac{4}{3}}^{(\infty)}$ gives the asymptotic distillation rate, $\mathcal{R}_{\frac{4}{3}}^{(\infty)} = \alpha_{\frac{4}{3}}^{(\infty)} = \frac{1}{4}$. We note that this is the only case of a depolarization channel (excluding the trivial identity channel) where the quantum switch achieves a nonzero distillation rate, which is found to be $\mathcal{R}_{\frac{4}{3}}^{(1)} = \frac{1}{3}$. With each subsequent super-switch, the distillation rate is decreasing, tending to the value $\mathcal{R}_{\frac{4}{3}}^{(\infty)}$ from above.

The general case corresponding to starting with a depolarization channel with any value $p \in (0, \frac{4}{3})$ is given by the solution

$$\begin{aligned}\alpha^{(\infty)} &= \frac{2 - \sqrt{3}}{4}, \quad \beta^{(\infty)} = \frac{\sqrt{3}}{4}, \quad \gamma^{(\infty)} = \frac{1}{2}, \\ p^{(\infty)} &= \frac{3 - \sqrt{3}}{6}.\end{aligned}\quad (\text{J14})$$

Thus, the asymptotic distillation rate for any depolarization channel with $p \neq 0, \frac{4}{3}$ is $\mathcal{R}_{\text{dep}}^{(\infty)} = \frac{(2 - \sqrt{3})}{4} \approx 0.067$. The fixed point solution tends to

$$\begin{aligned}\mathcal{S}^{(\infty)} &= \left(\frac{2 - \sqrt{3}}{4}\right) \mathbf{Id} \otimes \rho_{\text{Id}} + \left(\frac{\sqrt{3}}{4}\right) \mathcal{D}_{\frac{4}{3}} \otimes \rho_{\mathcal{N}} \\ &+ \frac{1}{2} \mathcal{D}_{\frac{3 - \sqrt{3}}{6}} \otimes \rho_{\mathcal{P}}.\end{aligned}\quad (\text{J15})$$

Interestingly, tracing out the ancilla we find that the fixed point corresponds to a completely depolarizing channel since in terms of the probability vector $\bar{p}$ we have

$$\begin{aligned}\mathcal{S}^{(\infty)} &= \left(\frac{2 - \sqrt{3}}{4}\right) \{1, 0, 0, 0\} + \left(\frac{\sqrt{3}}{4}\right) \left\{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right\} \\ &+ \frac{1}{2} \left\{1 - \frac{3p}{4}, \frac{p}{4}, \frac{p}{4}, \frac{p}{4}\right\} = \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}.\end{aligned}\quad (\text{J16})$$

The fixed points for the general case [Eq. (J10)] are identical to the case of the depolarization channel. Excluding points on the faces of the tetrahedron, we find that the only solution is $\bar{p}^{(\infty)} = \{1 - \frac{3p^{(\infty)}}{4}, \frac{p^{(\infty)}}{4}, \frac{p^{(\infty)}}{4}, \frac{p^{(\infty)}}{4}\}$ and $\bar{q} = \{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$ and the rest of the parameters identical to the one given by Eq. (J14). It turns out that in the general case, after a few iterations, the initial arbitrary vector $\bar{q}^{(n)}$ quickly tends to the vector $\{0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, while the vector $\bar{p}^{(n)}$ tends to a probability vector of a depolarization channel and thus the general case quickly converges to the case of the depolarization channels. Thus, the asymptotic distillation rate for any channel inside the tetrahedron is $\mathcal{R}_{\text{in}}^{(\infty)} = \frac{(2 - \sqrt{3})}{4} \approx 0.067$.

2. Channels on the faces of the tetrahedron

From the four faces of the tetrahedron, we split the discussion into two cases, since the recurrence relations slightly differ:

- (1) The three faces consisting of Pauli channels $\mathcal{P}_{\bar{p}}$ with $p_0 \neq 0$ and exactly one $p_i = 0, i \in \{1, 2, 3\}$.
- (2) The face consisting of Pauli channels $\mathcal{P}_{\bar{p}}$ with $p_0 = 0$ and $p_i \neq 0, \forall i \in \{1, 2, 3\}$, that is, the base of the tetrahedron.

a. The three faces consisting of Pauli channels $\mathcal{P}_{\bar{p}}$ with $p_0 \neq 0$ and exactly one $p_i = 0, i \in \{1, 2, 3\}$

Without loss of generality, we will consider the case of the face consisting of channels of the form $\mathcal{P}_{\bar{p}} = (1 - p - q)\rho + pX\rho X + qY\rho Y$. The derivation for the other two cases is identical.

In contrast to the case of channels inside the tetrahedron, channels on the faces can also lead to unitary channels after combining in the quantum switch. Thus, unitary contributions must also be tracked at each order. As a result, at order n we need to track the four terms that lead to different behaviors: the three terms that were tracked for channels inside the tetrahedron, as well as an extra term that tracks a unitary channel. Specifically, we have

- (1) Outcomes that lead to the identity channel $\mathbf{Id}$.
- (2) Outcomes that lead to the unitary channel $\mathcal{U}_Z = Z(\cdot)Z$.
- (3) Outcomes that lead to channels with zero portion of the identity, $\mathcal{M}_{\bar{q}^{(n)}}$, with $\bar{q}^{(n)} \equiv \{0, q_1^{(n)}, q_2^{(n)}, 0\}$.
- (4) Outcomes that lead to Pauli channels that do not belong to either of the three aforementioned classes, $\mathcal{P}_{\bar{p}^{(n)}}$, with $\bar{p}^{(n)} \equiv \{p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, 0\}$.

The total channel resulting at order n of the superswitches can be written as

$$\begin{aligned}\mathcal{S}^{(n)} &= \alpha^{(n)} \mathbf{Id} \otimes \rho_{\text{Id}} + \beta^{(n)} \mathcal{U}_Z \otimes \rho_{\mathcal{U}} \\ &+ \gamma^{(n)} \mathcal{P}_{\bar{p}^{(n)}} \otimes \rho_{\mathcal{P}} + \delta^{(n)} \mathcal{M}_{\bar{q}^{(n)}} \otimes \rho_{\mathcal{M}}.\end{aligned}\quad (\text{J17})$$

By $\rho_{\text{Id}}, \rho_{\mathcal{U}}, \rho_{\mathcal{P}}$, and $\rho_{\mathcal{M}}$, we denote the unnormalized states with orthogonal supports that gather the outcomes $\rho_{\text{Id}}, \rho_{\mathcal{U}}, \rho_{\mathcal{P}}$, and $\rho_{\mathcal{M}}$ of measurements on the ancillas that lead to the identity channel $\mathbf{Id}$, the unitary channel $\mathcal{U}_Z$, a general Pauli channel on the face $\mathcal{P}_{\bar{p}^{(n)}}$, with $\bar{p}^{(n)} \equiv \{p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, 0\}$, and a channel with zero portion of the identity $\mathcal{M}_{\bar{q}^{(n)}}$, with $\bar{q}^{(n)} \equiv \{0, q_1^{(n)}, q_2^{(n)}, 0\}$, respectively.

As in the case inside the tetrahedron, the next order $n + 1$ follows by forming all possible combinations of channels of order n from Eq. (3). Explicitly, we find

$$\mathcal{S}^{(n+1)} = \alpha^{(n+1)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n)} \mathcal{U}_Z \otimes \rho_{\mathcal{U}} + \gamma^{(n+1)} \mathcal{P}_{\vec{p}^{(n+1)}} \otimes \rho_{\mathcal{P}} + \delta^{(n+1)} \mathcal{M}_{\vec{q}^{(n+1)}} \otimes \rho_{\mathcal{M}}, \quad (\text{J18})$$

where

$$\begin{aligned} \alpha^{(n+1)} &= (\alpha^{(n)})^2 + (\beta^{(n)})^2 + (\delta^{(n)})^2 (1 - 2q_1^{(n)} q_2^{(n)}), \\ \beta^{(n+1)} &= 2(\gamma^{(n)})^2 p_1^{(n)} p_2^{(n)} + 2(\delta^{(n)})^2 q_1^{(n)} q_2^{(n)} \\ &\quad + 2\gamma^{(n)} \delta^{(n)} (q_1^{(n)} p_2^{(n)} + p_1^{(n)} q_2^{(n)}) + 2\alpha^{(n)} \beta^{(n)}, \\ \gamma^{(n+1)} &= 2\alpha^{(n)} \gamma^{(n)} + (\gamma^{(n)})^2 (1 - 2p_1^{(n)} p_2^{(n)}) \\ &\quad + 2\gamma^{(n)} \delta^{(n)} (1 - q_1^{(n)} p_2^{(n)} - p_1^{(n)} q_2^{(n)}), \\ \delta^{(n+1)} &= 2\alpha^{(n)} \delta^{(n)} + 2\beta^{(n)} \delta^{(n)} + 2\beta^{(n)} \gamma^{(n)} (1 - p_0^{(n)}), \\ \vec{p}^{(n+1)} &= (2\gamma^{(n)} \delta^{(n)} \Pr(\mathbf{a}[\vec{p}^{(n)}, \vec{q}^{(n)}]) \vec{f}(\vec{p}^{(n)}, \vec{q}^{(n)}) \\ &\quad + (\gamma^{(n)})^2 \Pr(\mathbf{a}[\vec{p}^{(n)}, \vec{p}^{(n)}]) \vec{f}(\vec{p}^{(n)}, \vec{p}^{(n)}) \\ &\quad + 2\alpha^{(n)} \gamma^{(n)} \vec{p}^{(n)}) / \gamma^{(n+1)}, \\ \vec{q}^{(n+1)} &= (2\alpha^{(n)} \delta^{(n)} \vec{q}^{(n)} + 2\beta^{(n)} \gamma^{(n)} \{0, p_2^{(n)}, p_1^{(n)}, 0\} \\ &\quad + 2\beta^{(n)} \delta^{(n)} \{0, q_2^{(n)}, q_1^{(n)}, 0\}) / \delta^{(n+1)}. \end{aligned} \quad (\text{J19})$$

These are the recurrence relations that can be iterated to find the action of a superswitch of any order for channels at the face of the tetrahedron. The initial conditions for the recurrence relations are $\alpha^{(0)} = \beta^{(0)} = 0$, $\gamma^{(0)} = 1$, $\delta^{(0)} = 0$, $\vec{q}^{(0)} = 0$, and $\vec{p}^{(0)} = \vec{p}$. We note that in the case we are examining, the distillation rate at order n is given as the sum $\alpha^{(n)} + \beta^{(n)}$.

Looking for the fixed points in the interior of the edge, we find the only solution

$$\begin{aligned} \alpha^{(\infty)} &= \frac{2 - \sqrt{2}}{4}, \quad \beta^{(\infty)} = \gamma^{(\infty)} = \frac{1}{4}, \quad \delta^{(\infty)} = \frac{1}{2\sqrt{2}}, \\ \vec{p}^{(\infty)} &= \left\{ \sqrt{2} - 1, 1 - \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}}, 0 \right\}, \\ \vec{q}^{(\infty)} &= \left\{ 0, \frac{1}{2}, \frac{1}{2}, 0 \right\}, \end{aligned} \quad (\text{J20})$$

to which all points in the interior of the tetrahedron are attracted to. The asymptotic distillation rate is thus

$$\mathcal{R}_{\triangle}^{(\infty)} = \alpha^{(\infty)} + \beta^{(\infty)} = \frac{2 - \sqrt{2}}{4} + \frac{1}{4} \approx 0.3965. \quad (\text{J21})$$

b. The base of the tetrahedron, consisting of Pauli channels $\mathcal{P}_{\vec{p}}$ with $p_0 = 0$ and $p_i \neq 0, \forall i \in \{1, 2, 3\}$

The derivation of the last face, corresponding to channels of the form $\mathcal{P}_{\vec{p}} = pX\rho X + qY\rho Y + (1 - p - q)Z\rho Z$, follows the derivation of the other three faces with one difference: there is no possibility for unitary channel contributions. That is, the term $\beta^{(n)} \mathcal{U}_Z \otimes \rho_{\mathcal{U}}$ does not appear in Eq. (J17). Moreover, the general channel is a channel with zero portion of the identity and thus there is no $\mathcal{M}$ term; the superswitch

at order n will have the form

$$\mathcal{S}^{(n)} = \alpha^{(n)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n)} \mathcal{P}_{\vec{p}^{(n)}} \otimes \rho_{\mathcal{P}}. \quad (\text{J22})$$

At the next order, we will have

$$\mathcal{S}^{(n+1)} = \alpha^{(n+1)} \mathbf{Id} \otimes \rho_{\mathbf{Id}} + \beta^{(n+1)} \mathcal{P}_{\vec{p}^{(n+1)}} \otimes \rho_{\mathcal{P}}, \quad (\text{J23})$$

where

$$\begin{aligned} \alpha^{(n+1)} &= (\alpha^{(n)})^2 + (\beta^{(n)})^2 \\ &\quad \times (1 - 2q_1^{(n)} q_2^{(n)} - 2q_2^{(n)} q_3^{(n)} - 2q_3^{(n)} q_1^{(n)}), \\ \beta^{(n+1)} &= 2\alpha^{(n)} \beta^{(n)} + (\beta^{(n)})^2 \\ &\quad \times (2q_1^{(n)} q_2^{(n)} + 2q_2^{(n)} q_3^{(n)} + 2q_3^{(n)} q_1^{(n)}), \\ \vec{p}^{(n+1)} &= \frac{1}{\beta^{(n+1)}} (2\alpha^{(n)} \beta^{(n)} \vec{p}^{(n)} + 2(\beta^{(n)})^2 \\ &\quad \times \Pr(\mathbf{c}[\vec{p}^{(n)}, \vec{p}^{(n)}]) \vec{g}(\vec{p}^{(n)}, \vec{p}^{(n)})). \end{aligned} \quad (\text{J24})$$

The initial conditions in this case are $\alpha^{(0)} = 0$, $\beta^{(0)} = 1$, and $\vec{p}^{(0)} = \vec{p}$. Looking for the fixed points of the recurrence relations [Eq. (J24)], we find the solution

$$\alpha^{(\infty)} = \frac{1}{4}, \quad \beta^{(\infty)} = \frac{3}{4}, \quad \vec{p}^{(\infty)} = \left\{ 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right\}, \quad (\text{J25})$$

where we have restricted to points that are not in the edges. Thus, the distillation rate at the base of the tetrahedron is $\mathcal{R}_{\text{base}}^{(\infty)} = \frac{1}{4}$.

APPENDIX K: ASYMPTOTIC DISTILLATION RATES VERSUS FINITE-ORDER DISTILLATION RATES

The asymptotic distillation rates derived in Appendix J give the limiting values that can be achieved with SWITCH_n as the order n tends to infinity. For some channels, this implies that the higher the order of the superswitch, the higher the distillation rate. However, this is not true for all channels. In fact, there exist channels for which the highest distillation rate is achieved by the quantum switch, with each subsequent higher-order switch having a lower distillation rate, tending

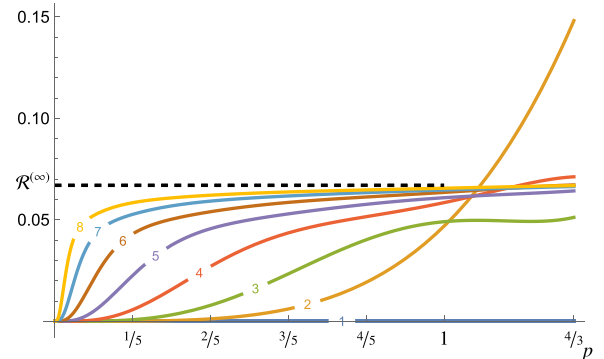


FIG. 8. The distillation rates $\mathcal{R}^{(n)}$ achieved by the superswitches from orders 1 to 8. The quantum switch cannot distill for any value of p , while each superswitch achieves a higher distillation rate for all values of $p \in [0, \frac{4}{3}]$, in comparison to all previous-order superswitches.

to the asymptotic rate from above. Channels that exhibit behavior in between the aforementioned two cases exist as well. This mirrors a similar phenomenon emerging in the problem of quantum state discrimination [21].

As a demonstrating example, we revisit the case of the depolarization channel $\mathcal{D}(\rho) = (1 - \frac{3p}{4})\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z)$ with $p \in [0, \frac{4}{3}]$. In Fig. 8, we plot distillation rates with SWITCH_n , $n \leq 8$.

We note that, for the depolarization channel, the point where the behavior changes coincides with the point where the Bloch vectors of the input states flip direction. That is, whenever $p > 1$ the Bloch vector $\vec{r} = (r_1, r_2, r_3)$ of a state ρ

is mapped to

$$\vec{r} = (r_1, r_2, r_3) \rightarrow (1 - p)(r_1, r_2, r_3), \quad (\text{K1})$$

and thus flips direction.

For general Pauli channels, although we do not have an exact characterization of the different regions, we have noticed by examining a number of examples that for channels close to the identity channel (1) the asymptotic rate bounds all finite orders from above and (2) a superswitch of order n will have a higher distillation rate than all superswitches of order m , with $m < n$.

-
- [1] L. Hardy, Towards quantum gravity: A framework for probabilistic theories with non-fixed causal structure, *J. Phys. A: Math. Theor.* **40**, 3081 (2007).
 - [2] L. Hardy, Quantum gravity computers: On the theory of computation with indefinite causal structure, in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*, edited by W. C. Myrvold and J. Christian (Springer Netherlands, Dordrecht, 2009), pp. 379–401.
 - [3] O. Oreshkov, F. Costa, and Č. Brukner, Quantum correlations with no causal order, *Nat. Commun.* **3**, 1092 (2012).
 - [4] G. Chiribella, G. M. D’Ariano, P. Perinotti, and B. Valiron, Quantum computations without definite causal structure, *Phys. Rev. A* **88**, 022318 (2013).
 - [5] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Transforming quantum operations: Quantum supermaps, *Europhys. Lett.* **83**, 30004 (2008).
 - [6] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Quantum circuit architecture, *Phys. Rev. Lett.* **101**, 060401 (2008).
 - [7] M. Araújo, F. Costa, and Č. Brukner, Computational advantage from quantum-controlled ordering of gates, *Phys. Rev. Lett.* **113**, 250402 (2014).
 - [8] M. M. Taddei, J. Cariñe, D. Martínez, T. García, N. Guerrero, A. A. Abbott, M. Araújo, C. Branciard, E. S. Gómez, S. P. Walborn, L. Aolita, and G. Lima, Computational advantage from the quantum superposition of multiple temporal orders of photonic gates, *PRX Quantum* **2**, 010320 (2021).
 - [9] X. Zhao, Y. Yang, and G. Chiribella, Quantum metrology with indefinite causal order, *Phys. Rev. Lett.* **124**, 190503 (2020).
 - [10] K. Simonov, G. Francica, G. Guarnieri, and M. Paternostro, Work extraction from coherently activated maps via quantum switch, *Phys. Rev. A* **105**, 032217 (2022).
 - [11] D. Ebler, S. Salek, and G. Chiribella, Enhanced communication with the assistance of indefinite causal order, *Phys. Rev. Lett.* **120**, 120502 (2018).
 - [12] S. Salek, D. Ebler, and G. Chiribella, Quantum communication in a superposition of causal orders, [arXiv:1809.06655](https://arxiv.org/abs/1809.06655) [quant-ph].
 - [13] F. Del Santo and B. Dakić, Two-Way communication with a single quantum particle, *Phys. Rev. Lett.* **120**, 060503 (2018).
 - [14] M. Caleffi and A. S. Cacciapuoti, Quantum switch for the quantum internet: Noiseless communications through noisy channels, *IEEE J. Sel. Areas Commun.* **38**, 575 (2020).
 - [15] L. M. Procopio, F. Delgado, M. Enríquez, N. Belabas, and J. A. Levenson, Communication enhancement through quantum coherent control of N channels in an indefinite causal-order scenario, *Entropy* **21**, 1012 (2019).
 - [16] L. M. Procopio, F. Delgado, M. Enríquez, N. Belabas, and J. A. Levenson, Sending classical information via three noisy channels in superposition of causal orders, *Phys. Rev. A* **101**, 012346 (2020).
 - [17] S. Sazim, M. Sedlak, K. Singh, and A. K. Pati, Classical communication with indefinite causal order for N completely depolarizing channels, *Phys. Rev. A* **103**, 062610 (2021).
 - [18] G. Chiribella, M. Wilson, and H. F. Chau, Quantum and classical data transmission through completely depolarizing channels in a superposition of cyclic orders, *Phys. Rev. Lett.* **127**, 190502 (2021).
 - [19] C. Mukhopadhyay and A. K. Pati, Superposition of causal order enables quantum advantage in teleportation under very noisy channels, *J. Phys. Commun.* **4**, 105003 (2020).
 - [20] D. Das and S. Bandyopadhyay, Quantum communication using a quantum switch of quantum switches, *Proc. R. Soc. A* **478**, 20220231 (2022).
 - [21] S. Kechrimparis, J. Moran, A. Karsa, C. Lee, and H. Kwon, Enhancing quantum state discrimination with indefinite causal order, *New J. Phys.* **26**, 123030 (2024).
 - [22] L. A. Rozema, T. Strömberg, H. Cao, Y. Guo, B.-H. Liu, and P. Walther, Experimental aspects of indefinite causal order in quantum mechanics, *Nat. Rev. Phys.* **6**, 483 (2024).
 - [23] G. Chiribella, M. Banik, S. S. Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal, and G. Kar, Indefinite causal order enables perfect quantum communication with zero capacity channels, *New J. Phys.* **23**, 033039 (2021).
 - [24] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Purification of noisy entanglement and faithful teleportation via noisy channels, *Phys. Rev. Lett.* **76**, 722 (1996).
 - [25] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, *Phys. Rev. A* **53**, 2046 (1996).
 - [26] A. Kent, Entangled mixed states and local purification, *Phys. Rev. Lett.* **81**, 2839 (1998).
 - [27] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, *Phys. Rev. A* **60**, 1888 (1999).

- [28] H.-K. Lo and S. Popescu, Concentrating entanglement by local actions: Beyond mean values, *Phys. Rev. A* **63**, 022301 (2001).
- [29] E. Jané, Purification of two-qubit mixed states, *Quantum Inf. Comput.* **2**, 348 (2002).
- [30] P. Horodecki and M. Demianowicz, Fidelity thresholds in single copy entanglement distillation, *Phys. Lett. A* **354**, 40 (2006).
- [31] B. Regula, K. Fang, X. Wang, and M. Gu, One-shot entanglement distillation beyond local operations and classical communication, *New J. Phys.* **21**, 103017 (2019).
- [32] A. Streltsov, G. Adesso, and M. B. Plenio, *Colloquium*: Quantum coherence as a resource, *Rev. Mod. Phys.* **89**, 041003 (2017).
- [33] K.-D. Wu, T. Theurer, G.-Y. Xiang, C.-F. Li, G.-C. Guo, M. B. Plenio, and A. Streltsov, Quantum coherence and state conversion: Theory and experiment, *npj Quantum Inf.* **6**, 22 (2020).
- [34] C. L. Liu and C. P. Sun, Optimal probabilistic distillation of quantum coherence, *Phys. Rev. Res.* **3**, 043220 (2021).
- [35] M. Forster, S. Winkler, and S. Wolf, Distilling nonlocality, *Phys. Rev. Lett.* **102**, 120401 (2009).
- [36] G. Eftaxias, M. Weilenmann, and R. Colbeck, Advantages of multicopy nonlocality distillation and its application to minimizing communication complexity, *Phys. Rev. Lett.* **130**, 100201 (2023).
- [37] S. G. Naik, G. L. Sidhardh, S. Sen, A. Roy, A. Rai, and M. Banik, Distilling nonlocality in quantum correlations, *Phys. Rev. Lett.* **130**, 220201 (2023).
- [38] G. Chiribella and H. Kristjánsson, Quantum Shannon theory with superpositions of trajectories, *Proc. R. Soc. A* **475**, 20180903 (2019).
- [39] M.-D. Choi, Completely positive linear maps on complex matrices, *Linear Algebra Appl.* **10**, 285 (1975).
- [40] A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Rep. Math. Phys.* **3**, 275 (1972).
- [41] A. Mitra, H. Badhani, and S. Ghosh, Improvement in quantum communication using quantum switch, *Phys. Scr.* **98**, 045101 (2023).
- [42] Z. Wu, J. Fullwood, Z. Ma, S.-Q. Zhou, Q. Zhao, and G. Chiribella, General communication enhancement via the quantum switch, *Phys. Rev. A* **111**, 012605 (2025).
- [43] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
- [44] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *J. Math. Phys.* **48**, 052104 (2007).
- [45] N. J. A. Sloane, A library of Hadamard matrices, <http://neilsloane.com/hadamard/>.
- [46] J. H. Dinitz and D. R. Stinson, *Contemporary Design Theory: A Collection of Surveys* (Wiley, New York, 1992).
- [47] M. B. Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M_2 , *Linear Algebra Appl.* **347**, 159 (2002).
- [48] D. Davalos, M. Ziman, and C. Pineda, Divisibility of qubit channels and dynamical maps, *Quantum* **3**, 144 (2019).
- [49] M. Horodecki, P. W. Shor, and M. B. Ruskai, Entanglement breaking channels, *Rev. Math. Phys.* **15**, 629 (2003).
- [50] T. S. Cubitt, M. B. Ruskai, and G. Smith, The structure of degradable quantum channels, *J. Math. Phys.* **49**, 102104 (2008).
- [51] C.-S. Niu and R. B. Griffiths, Optimal copying of one quantum bit, *Phys. Rev. A* **58**, 4377 (1998).
- [52] N. J. Cerf, Pauli cloning of a quantum bit, *Phys. Rev. Lett.* **84**, 4497 (2000).
- [53] M. Jiang, S. Luo, and S. Fu, Channel-state duality, *Phys. Rev. A* **87**, 022310 (2013).
- [54] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, *Phys. Rev. A* **40**, 4277 (1989).
- [55] H. Ko, S. Kechrimparis, Y.-H. Ko, K.-J. Kim, B.-S. Choi, C. M. Kropf, C. J. Youn, and J. Bae, Secure quantum communication with the preservation of optimal measurements, *IEEE J. Sel. Areas Commun.* **43**, 2798 (2025).