Supplementary Materials for "Approximate Autonomous Quantum Error Correction with Reinforcement Learning"

Yexiong Zeng,[1,2] Zheng-Yang Zhou,[1] Enrico Rinaldi,[3,1,2,4,5] Clemens Gneiting,[1,2,*] and Franco Nori[1,2,4,†]

[1] *Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan*
[2] *Quantum Computing Center, RIKEN, Wakoshi, Saitama 351-0198, Japan*
[3] *Quantinuum K.K., Otemachi Financial City Grand Cube 3F, 1-9-2 Otemachi, Chiyoda-ku, Tokyo, Japan*
[4] *Department of Physics, The University of Michigan, Ann Arbor, MI, 48109-1040, USA*
[5] *Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Wakoshi, Saitama 351-0198, Japan*

(Dated: July 6, 2023)

CONTENTS

S1. MEAN FIDELITY OF THE CODE SPACE

In this section, we determine the fidelity between initial and evolved code states averaged over the code space. An arbitrary state in the code space (i.e., the initial state) can be expanded as

$$|\psi_{\theta\phi}\rangle = \cos\frac{\theta}{2}|0_{\rm L}\rangle + e^{i\phi}\sin\frac{\theta}{2}|1_{\rm L}\rangle\,, \tag{1}$$

where $|0_{\rm L}\rangle$, $|1_{\rm L}\rangle$ are the logical codewords, and θ, ϕ are the familiar Bloch sphere angles. A linear, positive definite, and trace-preserving map $\mathcal{M}[\cdot]$ evolves the initial state $\rho_{t_0}(\theta,\phi) = |\psi_{\theta\phi}\rangle\langle\psi_{\theta\phi}|$ to the quantum state $\rho_t(\theta,\phi) = \mathcal{M}[\rho_{t_0}(\theta,\phi)]$ at time t. With the fidelity between the evolved state $\rho_t(\theta,\phi)$ and the initial state $\rho_{t_0}(\theta,\phi)$ given by $F(\theta,\phi,t) = {\rm Tr}[\rho_{t_0}(\theta,\phi)\rho_t(\theta,\phi)]$, the mean fidelity of the code space can be written as

$$\bar{F}(t) = \frac{1}{4\pi}\int_\Omega F(\theta,\phi,t)\ {\rm d}\Omega. \tag{2}$$

Next, we expand the initial state in terms of the logical Pauli operators

$$\rho_{t_0}(\theta,\phi) = |\psi_{\theta,\phi}\rangle\langle\psi_{\theta,\phi}| = \sum_{j=0,x,y,z} c_j(\theta,\phi)\frac{\sigma_j}{2}, \tag{3}$$

where

$$\begin{aligned} \sigma_0 &= |0_{\rm L}\rangle\langle 0_{\rm L}| + |1_{\rm L}\rangle\langle 1_{\rm L}|, \qquad \sigma_x = |0_{\rm L}\rangle\langle 1_{\rm L}| + |1_{\rm L}\rangle\langle 0_{\rm L}|, \\ \sigma_y &= i(|0_{\rm L}\rangle\langle 1_{\rm L}| - |1_{\rm L}\rangle\langle 0_{\rm L}|), \quad \sigma_z = |1_{\rm L}\rangle\langle 1_{\rm L}| - |0_{\rm L}\rangle\langle 0_{\rm L}|, \end{aligned} \tag{4}$$

* clemens.gneiting@riken.jp
† fnori@riken.jp

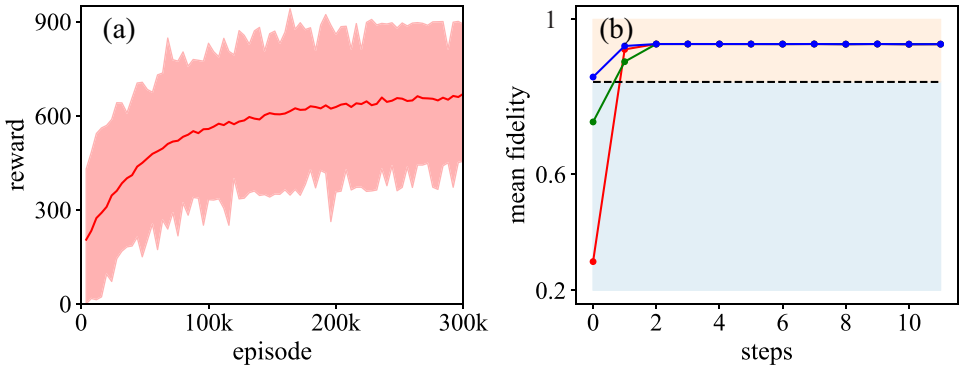

FIG. S1. (a) Plot of the maximum (upper bound of the shaded area), average (solid red line), and minimum rewards (lower bound of the shaded area) as functions of the episodes during the training process, where the dynamic evolution time is $\gamma_a t = 0.6$. (b) Evolution of the mean fidelity with the number of steps for three randomly selected episodes. Other parameters are $g/\gamma_a = 400$ and $\gamma_b/\gamma_a = 1750$ (i.e., $C \approx 91.4$).

and the elements of the coherent Bloch vector are

$$c_0(\theta,\phi) = 1, \quad c_x(\theta,\phi) = \sin\theta\cos\phi, \quad c_y(\theta,\phi) = \sin\theta\sin\phi, \quad c_z(\theta,\phi) = \cos\theta. \tag{5}$$

By combining Eqs. (2), (3), (4), (5), we can reformulate the mean fidelity as

$$\begin{aligned}\bar{F}(t) &= \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} \mathrm{Tr}\left(\sum_j c_j(\theta,\phi)\frac{\sigma_j}{2}\mathcal{M}\left[\sum_k c_k(\theta,\phi)\frac{\sigma_k}{2}\right]\right)\sin\theta\ \mathrm{d}\phi\mathrm{d}\theta \\ &= \sum_{jk}\frac{1}{4\pi}\int_\theta\int_\phi c_j c_k \sin\theta\ \mathrm{d}\phi\mathrm{d}\theta\,\mathrm{Tr}\left(\frac{\sigma_j}{2}\mathcal{M}\left[\frac{\sigma_k}{2}\right]\right) \\ &= \frac{1}{6}\sum_{j=\pm x,\pm y,\pm z}\mathrm{Tr}\left(\rho_j\mathcal{M}\left[\rho_j\right]\right),\end{aligned} \tag{6}$$

where we have defined $\rho_0 = \sigma_0/2$ and $\rho_{\pm j} = (\sigma_0 \pm \sigma_j)/2$ $(j = x, y, z)$. This result is consistent with previous work [1].

S2. SEARCH FOR THE OPTIMAL CODE SPACE WITH REINFORCEMENT LEARNING

Finding the optimal coefficients $c_n^{(0)}$ and $c_n^{(1)}$, such that the codewords, Eq. (2) of the main text, maximize the mean fidelity $\bar{F}(|0_\mathrm{L}\rangle, |1_\mathrm{L}\rangle)$ at some fixed reference time, represents a complex optimization problem that we solve using reinforcement learning (RL). In brief, we divide each episode into a finite number of steps $k = 1, 2, \cdots, K$. At step k the agent observes the state $s_k \in S$ of the surrounding environment and chooses an action a_k according to the policy $\pi(A|S)$. Then the agent obtains the new state s_{k+1}, and the environment returns a reward r_{k+1}. The policy $\pi(A|S)$ is updated via the experience data to maximize the accumulated reward R. The state s_k contains six observations $\mathrm{Tr}(\rho_j\mathcal{M}(\rho_j))$, $j = \pm x, \pm y, \pm z$ [see Eq.(6)] that can be calculated by using the toolkit Qutip [2, 3]. The action of the agent consists of the vector of coefficients $[c_n^{(0)}, c_n^{(1)}]$. The reward r_{k+1} is proportional to the difference between the mean fidelities of the code space obtained by the policy function and of the break-even point, $\epsilon_k = \bar{F}_k(|0_L\rangle, |1_L\rangle) - \bar{F}(|0\rangle, |1\rangle)$. The specific reward scheme is as follows: if $\epsilon_{k+1} > 0$ and $\epsilon_{k+1} > \epsilon_k$, the reward is $1000\epsilon_{k+1}$; if $\epsilon_{k+1} > 0$ and $\epsilon_{k+1} < \epsilon_k$, the reward is $100\epsilon_{k+1}$; if $\epsilon_k \leq 0$, the reward is 0.

We truncate the code space to 6 photons and set the maximum number of steps per episode to $K = 11$. The parameters for the simulation of the dynamics are chosen as $g/\gamma_a = 400$, $\gamma_b/\gamma_a = 1750$ (i.e., $C \approx 91.4$), $\gamma_a = 0.02$ MHz, and $\gamma_a t = 0.6$, in agreement with current experimental conditions. We apply the proximal policy optimization algorithm to optimize the policy [4], which is achieved by using the Python toolkit Ray [5]. The parameters of the neural networks are the default ones offered by Ray.

As shown in Fig. S1(a), the mean reward approximately converges to a constant after about 200k episodes. Meanwhile, the max reward is obtained when the coefficient is $c_1^{(0)} \approx 1$, $c_0^{(1)} \approx 1$, and $c_0^{(0)} \approx c_1^{(1)} \approx 0$. We thus conclude that the optimal code space consists of the Fock states $|2\rangle$ and $|4\rangle$.

To demonstrate that different initial states converge to the same mean fidelity, we display in Fig. S1(b) three random episodes. One can easily see that all three episodes converge towards the same mean fidelity (well above the break-even point of 0.84), in each case yielding the same optimal codewords, the Fock states $|2\rangle$ and $|4\rangle$, as described by the RL code.

S3. ASSESSING THE PERFORMANCE OF THE APPROXIMATE AQEC

Here we analytically assess the performance of the approximate AQEC. First, we solve the dynamical evolution for an arbitrary initial code state, and then calculate the mean fidelity. The dynamic evolution of the hybrid system composed of the encoding and the auxiliary mode is governed by the master equation

$$\frac{d\rho}{dt} = -i[H_I, \rho] + \frac{\gamma_a}{2}\mathcal{D}[a] + \frac{\gamma_b}{2}\mathcal{D}[\sigma_-], \tag{7}$$

where the Hamiltonian of the hybrid system is $H_I = g(L_{\rm eng}\sigma_+ + L_{\rm eng}^\dagger\sigma_-)$, and the resulting error correction operator is the same as in the main text.

If the parameters satisfy the conditions $g, \gamma_a \ll \gamma_b$, we can approximately write the density operator of the hybrid system as $\rho(t) = \rho_a(t) \otimes |0\rangle\langle 0|$, where $\rho_a(t)$ is the state of the encoding mode and $|0\rangle\langle 0|$ is the ground state of the auxiliary qubit. Therefore, the dynamical evolution of the encoding mode is governed by the effective master equation [6]

$$\frac{d\rho_a}{dt} = \frac{\gamma_a}{2}\mathcal{D}[a] + \frac{\gamma_a\lambda}{2}\mathcal{D}[L_{\rm eng}], \quad \lambda = \frac{8|g|^2}{\gamma_b\gamma_a} = 8C. \tag{8}$$

Without loss of generality, we can restrict the effective master equation to the first five Fock states (from the ground state to four photons). The density matrix elements then satisfy the equations

$$\begin{aligned}
&\frac{1}{\gamma_a}\frac{d\rho_{22}}{dt} = \frac{1}{2}(6\rho_{33} - 4\rho_{22} + \lambda\rho_{11}), && \frac{1}{\gamma_a}\frac{d\rho_{24}}{dt} = \frac{1}{2}(-6\rho_{24} + \lambda\rho_{13}),\\
&\frac{1}{\gamma_a}\frac{d\rho_{44}}{dt} = \frac{1}{2}(-8\rho_{44} + \lambda\rho_{33}), && \frac{1}{\gamma_a}\frac{d\rho_{42}}{dt} = \frac{1}{2}(-6\rho_{42} + \lambda\rho_{31}),\\
&\frac{1}{\gamma_a}\frac{d\rho_{11}}{dt} = \frac{1}{2}(4\rho_{22} - 2\rho_{11} - \lambda\rho_{11}), && \frac{1}{\gamma_a}\frac{d\rho_{13}}{dt} = \frac{1}{2}(2\sqrt{8}\rho_{24} - 4\rho_{13} - \lambda\rho_{13}),\\
&\frac{1}{\gamma_a}\frac{d\rho_{33}}{dt} = \frac{1}{2}(8\rho_{44} - 6\rho_{33} - \lambda\rho_{33}), && \frac{1}{\gamma_a}\frac{d\rho_{31}}{dt} = \frac{1}{2}(2\sqrt{8}\rho_{42} - 4\rho_{31} - \lambda\rho_{31}),
\end{aligned}$$

where we have defined the density matrix as $\rho_a(t) = \sum_{ij}\rho_{ij}(t)|i\rangle\langle j|$. If we assume $\lambda \gg 1$ and $\lambda \gg \gamma_a t$, we obtain the approximate solution

$$\begin{aligned}
\rho_{00} \approx& 1 - \left(\frac{6}{5} + \frac{48}{25\lambda}\right)\rho_{44}(0)\left\{\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&- \rho_{22}(0)\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \rho_{44}(0)\exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{11} \approx& \frac{24}{5\lambda}\rho_{44}(0)\left\{\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&+ \frac{4}{\lambda}\rho_{22}(0)\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{22} \approx& \left(\frac{6}{5} - \frac{72}{25\lambda}\right)\rho_{44}(0)\left\{\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&+ \left(1 - \frac{4}{\lambda}\right)\rho_{22}(0)\exp\left[-\frac{4}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{33} \approx& \frac{8}{\lambda}\rho_{44}(0)\exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,
\end{aligned}$$

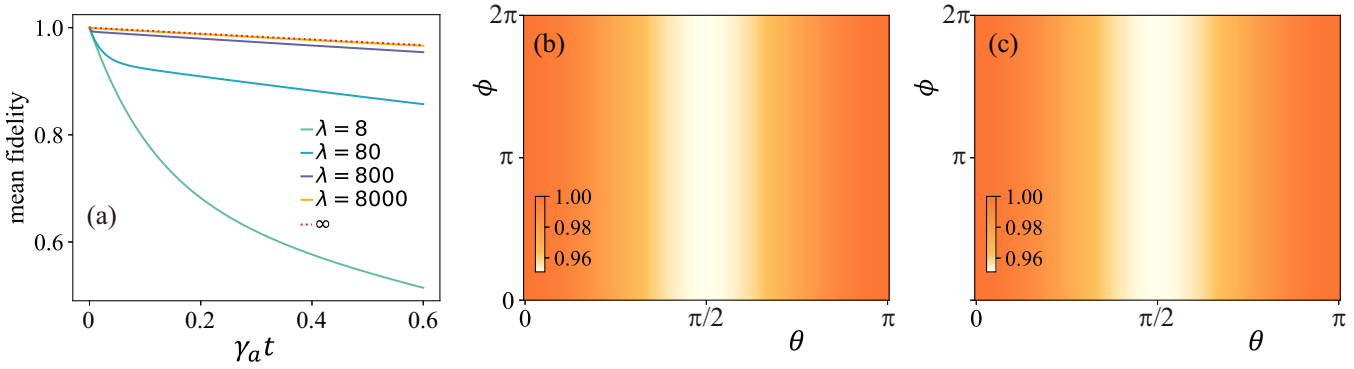

FIG. S2. (a) Evolution of the mean fidelity for different choices of λ. (b) The analytic fidelity $F(\rho_a(\theta,\phi,0),\rho_a(\theta,\phi,t))$ (i.e., Eq.(12)) as function of the Bloch angles θ and ϕ. (c) The numerical fidelity $F(\rho_a(\theta,\phi,0),\rho_a(\theta,\phi,t))$ (i.e., numerically simulating the master equation Eq.(8) with $\lambda = 50,000$) as function of the Bloch angles θ and ϕ. Other parameters are $\gamma_a t = 0.6$.

$$
\begin{aligned}
\rho_{44} &\approx \left(1-\frac{8}{\lambda}\right)\rho_{44}(0)\exp\left[-\frac{24}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{24} &\approx \left(1-\frac{4\sqrt{2}}{\lambda}\right)\rho_{24}(0)\exp\left[\frac{4(\sqrt{2}-4)\gamma_a t}{\lambda} + (2\sqrt{2}-3)\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{13} &\approx \frac{4\sqrt{2}}{\lambda}\rho_{24}(0)\exp\left[\frac{4(\sqrt{2}-4)\gamma_a t}{\lambda} + (2\sqrt{2}-3)\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,
\end{aligned}
\tag{9}
$$

where we have expanded the elements of the density matrix to the first order of $\frac{1}{\lambda}$ and $\frac{\gamma_a t}{\lambda}$. Similar to other autonomous error correction schemes [1, 7], the error correction efficiency based on the RL code also depends on the evaluation time $\gamma_a t$. If $\gamma_a t$ is too large, the system has decayed into the ground state ($\rho_{00}(t) \approx 1$), i.e., the capability for AQEC is lost. Therefore, we limit $\gamma_a t$ to an appropriate scale $\gamma_a t \leq 0.6$. If $\lambda \gg 24$ (i.e., $\lambda \to \infty$), we can approximate the density matrix to order zero in the parameters $\frac{1}{\lambda}$ and $\frac{\gamma_a t}{\lambda}$

$$
\rho_a(t) \approx \begin{pmatrix}
0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & \rho_{22}(0) & 0 & \rho_{24}(0)\exp(-u\gamma_a t)\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & \rho_{42}(0)\exp(-u\gamma_a t) & 0 & \rho_{44}(0)
\end{pmatrix}, \quad u = 3-2\sqrt{2} \approx 0.17. \tag{10}
$$

With this, we can derive the approximate mean fidelity of the RL code space

$$
\begin{aligned}
\bar{F}(t) &= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left[\rho_{22}^2(0)+\rho_{44}^2(0)+2|\rho_{24}(0)|^2\exp(-u\gamma_a t)\right]\sin\theta\ \mathrm{d}\theta\mathrm{d}\phi\\
&= \frac{1}{2}\int_0^{\pi}\left\{1+2\sin^2\left(\frac{\theta}{2}\right)\cos^2\left(\frac{\theta}{2}\right)\left[\exp(-u\gamma_a t)-1\right]\right\}\sin\theta\ \mathrm{d}\theta\\
&= \frac{2}{3}+\frac{1}{3}\exp(-u\gamma_a t),
\end{aligned}
\tag{11}
$$

which can be further simplified as $\bar{F}(t) = 1-\frac{1}{3}u\gamma_a t$ due to the condition $u\gamma_a t \ll 1$. Moreover, the fidelity between an individual initial state $\rho_a(\theta,\phi,0)$ and the corresponding evolved state $\rho_a(\theta,\phi,t)$ is

$$
F(\rho_a(\theta,\phi,0),\rho_a(\theta,\phi,t)) = \mathrm{Tr}[|\psi_{\theta\phi}\rangle\langle\psi_{\theta\phi}|\,\rho(\theta,\phi,t)] = 1+2\sin^2\left(\frac{\theta}{2}\right)\cos^2\left(\frac{\theta}{2}\right)\left[\exp(-u\gamma_a t)-1\right], \tag{12}
$$

which depends on the Bloch angle θ, but not on the Bloch angle ϕ for $\lambda \to \infty$. If the angle θ is equal to 0 and π, the fidelity is close to the maximum value of 1; if it is $\frac{\pi}{2}$, then the fidelity is the lowest (about 0.95), but still well above the break-even point (about 0.84). For a comparison between the numerical and analytical results, we simulate the

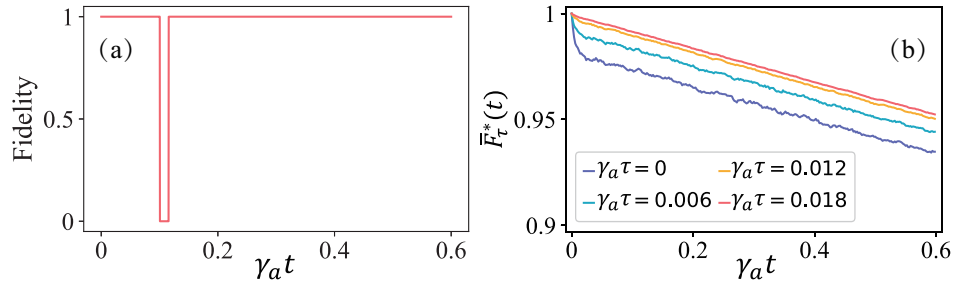

FIG. S3. Performance of the RL code in terms of the error-correctable fidelity. (a) Fidelity $\langle 2|\rho(t)|2\rangle$ for a random quantum trajectory with the initial state $|2\rangle$. (b) Trajectory-wise, temporally coarse-grained fidelity $\bar{F}^*_\tau(t)$, averaged over the code space and over 10^4 quantum trajectories, with $\gamma_a\tau = 0,\ 0.006,\ 0.012,$ and 0.018; other parameters are the same as in Fig. S1. For sufficiently large τ, the initial, delay-induced dip is removed. The resulting error-correctable fidelity reflects the irreversible fidelity loss.

time-dependent mean fidelity with different λ in Fig. S2(a). Our results demonstrate that the mean fidelity increases with increasing λ. Moreover, the numerical mean fidelity (i.e., simulating the master Eq. (8) with $\lambda = 8000$) well agrees with the analytical solution Eq. (11) (i.e., $\lambda = \infty$). Similarly, we get very good agreement between analytical prediction and numerical evaluation for individual state fidelities, as shown in Figs. S2(b) and (c). Numerical fidelities are obtained by simulating the master Eq. (8) with $\lambda = 50,000$. This shows that our analytical results are reliable for $\lambda \gg 24$.

S4. ERROR-CORRECTABLE FIDELITY

As a consequence of the finite rate of the engineered jump operator in AQEC, there is a delay between the occurrence of an error and the onset of the recovery jump. This becomes transparent if one unravels the evolution under the master equation (7) in terms of individual quantum trajectories. In Fig. S3(a) we demonstrate this with a random quantum trajectory for the initial state $|2\rangle$. Note that the fidelity drops to zero when an error occurs, and recovers to close to unity once the delayed recovery jump occurs (Let us clarify that other initial states, specifically, superpositions of the code words, would recover to fidelity values close to unity but lower, reflecting the approximate nature of the AQEC). Therefore, this random delay reduces the average fidelity when averaged over many trajectories, resulting in an initial dip as observed in Fig. S3(b) for $\gamma_a\tau = 0$. However, the associated fidelity loss is, in principle, recoverable, as the information, while stored in the error space, is not lost. It is therefore instructive to distill the irreversible part of the fidelity loss, caused by the occurrence of multiple errors before correction and the imperfect state recovery under approximate AQEC. This can be achieved by using a temporally coarse-grained redefinition of the fidelity [8]:

$$F^*_\tau(t) = \max_{t^*\in[t,\ t+\tau]} F(t^*). \tag{13}$$

If the coarse-graining parameter τ is chosen sufficiently large, i.e., on the order of the average time delay, then the fidelity (13) ignores the delay-induced fidelity loss, reflecting the error-correctable fidelity. In Fig. S3(b) we demonstrate this with the RL code for different choices of τ. As expected, we find that the initial dip of the fidelity is removed with increasing τ.

While the error-correctable fidelity provides valuable insight about the in-principle achievable performance of the autonomously corrected RL code, the standard fidelity remains the operationally relevant one. This is because, in the absence of monitoring, that is, if no additional information is gained about whether the state resides in the code space or in the error space, gate operations must be conducted under the (due to the delayed-error-correction process not necessarily correct) assumption that, at the time of the gate operation, the state resides in the code space.

S5. NAIVE ERROR CORRECTION OPERATOR

Let us consider another jump operator that may appear to counteract well the single-photon loss. When a single-photon loss occurs, the encoded state is changed to the error state

$$a(\cos\frac{\theta}{2}|2\rangle + e^{i\phi}\sin\frac{\theta}{2}|4\rangle) \quad\longrightarrow\quad \cos\frac{\theta}{2}\sqrt{2}|1\rangle + \sqrt{4}e^{i\phi}\sin\frac{\theta}{2}|3\rangle. \tag{14}$$

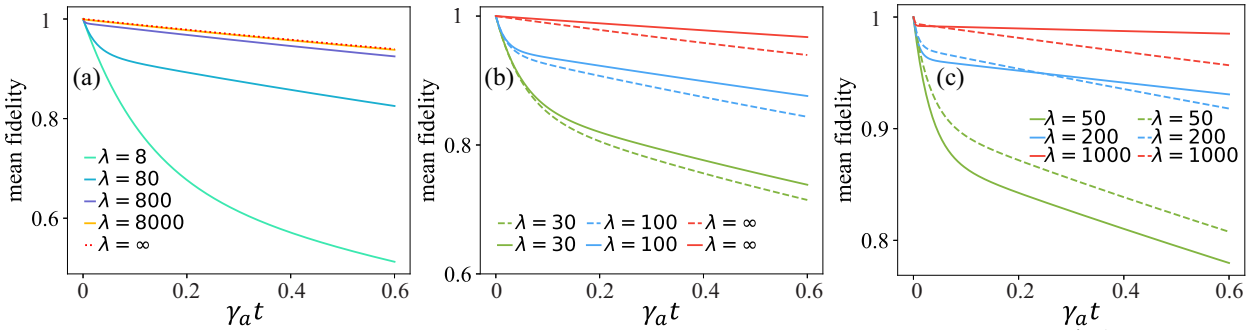

FIG. S4. (a) Mean fidelity versus time for different λ (the red dotted line describes the analytical result); (b) Comparison of the error correction performance of the error correction operator Eq. (15) (dashed line) and the RL error correction operator from the main text (solid line); (c) Evolution of the mean fidelity resulting from the original master equation Eq. (8) (dashed line) or from the modified master equation ($a \to a + a_1$) (solid line) for different λ.

To steer the error state back to the code space undisturbed, one may tentatively use the following error correction operator, which is different from the one in the main text:

$$L_{\rm o} = \sqrt{2}|2\rangle\langle 1| + |4\rangle\langle 3|, \quad L_{\rm eng} = \frac{L_{\rm o}}{\sqrt{{\rm Tr}[L_{\rm o}^\dagger L_{\rm o}]}}. \tag{15}$$

This jump operator revises the error state as

$$L_{\rm eng}\left(\cos\frac{\theta}{2}\sqrt{2}|1\rangle + \sqrt{4}e^{i\phi}\sin\frac{\theta}{2}|3\rangle\right) \quad \longrightarrow \quad \frac{2}{\sqrt{{\rm Tr}[L_{\rm o}^\dagger L_{\rm o}]}}\left(\cos\frac{\theta}{2}|2\rangle + e^{i\phi}\sin\frac{\theta}{2}|4\rangle\right), \tag{16}$$

that is, the encoded information remains undisturbed. To further investigate the efficiency of this error correction operator, we substitute it into the master Eq. (8) and restrict the system again to the first four photons, which yields the equations

$$\begin{aligned}
\frac{1}{\gamma_a}\frac{d\rho_{22}}{dt} &= \frac{1}{2}(6\rho_{33} - 4\rho_{22} + \frac{4}{3}\lambda\rho_{11}), & \frac{1}{\gamma_a}\frac{d\rho_{24}}{dt} &= \frac{1}{2}(-6\rho_{24} + \frac{2\sqrt{2}}{3}\lambda\rho_{13}),\\
\frac{1}{\gamma_a}\frac{d\rho_{44}}{dt} &= \frac{1}{2}(-8\rho_{44} + \frac{2}{3}\lambda\rho_{33}), & \frac{1}{\gamma_a}\frac{d\rho_{42}}{dt} &= \frac{1}{2}(-6\rho_{42} + \frac{2\sqrt{2}}{3}\lambda\rho_{31}),\\
\frac{1}{\gamma_a}\frac{d\rho_{11}}{dt} &= \frac{1}{2}(4\rho_{22} - 2\rho_{11} - \frac{4}{3}\lambda\rho_{11}), & \frac{1}{\gamma_a}\frac{d\rho_{13}}{dt} &= \frac{1}{2}(2\sqrt{8}\rho_{24} - 4\rho_{13} - \lambda\rho_{13}),\\
\frac{1}{\gamma_a}\frac{d\rho_{33}}{dt} &= \frac{1}{2}(8\rho_{44} - 6\rho_{33} - \frac{2}{3}\lambda\rho_{33}), & \frac{1}{\gamma_a}\frac{d\rho_{31}}{dt} &= \frac{1}{2}(2\sqrt{8}\rho_{42} - 4\rho_{31} - \lambda\rho_{31}).
\end{aligned} \tag{17}$$

In the limit $\lambda \to \infty$, we obtain an approximate density matrix, which has the same form as Eq.(10), except for that $u = \frac{1}{3}$. Correspondingly, the expressions for the fidelities $F(|\psi_{\theta\phi}\rangle\langle\psi_{\theta\phi}|, \rho_a(\theta,\phi,t))$ and $\bar{F}(t)$ remain the same, cf. Eqs. (11) and (12): the only difference is $u = \frac{1}{3} > 3 - 2\sqrt{2}$. In Fig. S4(a), we show the fidelity over time for different λ. We find that the analytical results coincide well with the numerical results for large enough λ. In Fig. S4(b), we see that the efficiency of the RL error correction operator from the main text is higher than the error correction operator Eq. (15). Why is this? We can provide a physical reason: although the error correction operator Eq. (15) ensures that the information is not disturbed when the jumps occur, the time evolution in between the jumps, captured by the non-Hermitian terms in the master equation, acts detrimentally on the encoded information (i.e., $\lambda\langle 1|F_{\rm eng}^\dagger F_{\rm eng}|1\rangle \gg \lambda\langle 3|F_{\rm eng}^\dagger F_{\rm eng}|3\rangle$), resulting in an overall reduced performance compared to the RL error correction operator of the main text.

S6. ANALYSIS OF THE KNILL-LAFLAMME CONDITION

The mean fidelity of the RL code, Eq. (11) remains slightly below one because the code words have different mean photon numbers: the Knill-Laflamme (KL) condition is only partially satisfied [9]. To understand the consequences

of this better, let us modify the master equation by replacing the single-photon loss jump operator a by $a + a_1$, where $a_1 = (2 - \sqrt{2})|1\rangle\langle 2|$ (i.e., $\mathcal{D}[a] \to \mathcal{D}[a + a_1]$). The modified jump operator now satisfies the condition $\langle 0_L|(a + a_1)^\dagger(a + a_1)|0_L\rangle = \langle 1_L|(a + a_1)^\dagger(a + a_1)|1_L\rangle$ for the RL code, that is, the KL condition is satisfied. Therefore, we can completely correct the error $a + a_1$ under the conditions $\frac{8g^2}{\gamma_a\gamma_b} \to \infty$, $\frac{1}{\lambda} \ll 1$, and $\frac{\gamma_a t}{\lambda} \ll 1$. Next, we show that the mean fidelity can indeed reach unity under the 0-order approximation of the parameters $\frac{1}{\lambda}$ and $\frac{\gamma_a t}{\lambda}$. To this end, let us again evaluate the corresponding master equation for the system up to four photons,

$$\begin{aligned}
\frac{1}{\gamma_a}\frac{d\rho_{22}}{dt} &= \frac{1}{2}(6\rho_{33} - 8\rho_{22} + \lambda\rho_{11}), & \frac{1}{\gamma_a}\frac{d\rho_{24}}{dt} &= \frac{1}{2}(-8\rho_{24} + \lambda\rho_{13}),\\
\frac{1}{\gamma_a}\frac{d\rho_{44}}{dt} &= \frac{1}{2}(-8\rho_{44} + \lambda\rho_{33}), & \frac{1}{\gamma_a}\frac{d\rho_{42}}{dt} &= \frac{1}{2}(-8\rho_{42} + \lambda\rho_{31}),\\
\frac{1}{\gamma_a}\frac{d\rho_{11}}{dt} &= \frac{1}{2}(8\rho_{22} - 2\rho_{11} - \lambda\rho_{11}), & \frac{1}{\gamma_a}\frac{d\rho_{13}}{dt} &= \frac{1}{2}(8\rho_{24} - 4\rho_{13} - \lambda\rho_{13}),\\
\frac{1}{\gamma_a}\frac{d\rho_{33}}{dt} &= \frac{1}{2}(8\rho_{44} - 6\rho_{33} - \lambda\rho_{33}), & \frac{1}{\gamma_a}\frac{d\rho_{31}}{dt} &= \frac{1}{2}(8\rho_{42} - 4\rho_{31} - \lambda\rho_{31}).
\end{aligned} \tag{18}$$

We approximately solve this equation by expanding to first order in the parameters $\lambda \gg 1$ and $\lambda \gg \gamma_a t$,

$$\begin{aligned}
\rho_{00}(t) \approx& 1 - \left(\frac{3}{\lambda} + \frac{3}{2}\right)\rho_{44}(0)\left\{\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&- \rho_{44}(0)\exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \rho_{22}(0)\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{11}(t) \approx& \frac{12}{\lambda}\rho_{44}(0)\left\{\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&+ \frac{8}{\lambda}\rho_{22}(0)\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{22}(t) \approx& \left(\frac{3}{2} - \frac{9}{\lambda}\right)\rho_{44}(0)\left\{\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] - \exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right]\right\}\\
&+ \left(1 - \frac{8\gamma_a t}{\lambda}\right)\rho_{22}(0)\exp\left[-\frac{8\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{33}(t) \approx& \frac{8}{\lambda}\rho_{44}(0)\exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{44}(t) \approx& \left(1 - \frac{8}{\lambda}\right)\rho_{44}(0)\exp\left[-\frac{24\gamma_a t}{\lambda} + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{13}(t) \approx& \frac{8}{\lambda}\rho_{24}(0)\exp\left[-\frac{16}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2,\\
\rho_{24}(t) \approx& \left(1 - \frac{8}{\lambda}\right)\rho_{24}(0)\exp\left[-\frac{16}{\lambda}\gamma_a t + O\left(\frac{\gamma_a t}{\lambda}\right)^2\right] + O\left(\frac{1}{\lambda}\right)^2.
\end{aligned} \tag{19}$$

The density matrix is approximately equal to the initial state $\rho_a(0)$ for $\lambda \to \infty$, so the mean fidelity is approximately equal to unity. This implies that we can fully correct the single-photon loss by adding the small correction term a_1. In Fig. S4(c), we compare the error correction performance of the modified master equation ($a \to a + a_1$) and the original master equation (8) for different values of λ. With increasing λ, the modified master equation displays a better error correction ability. When λ is small, the mean fidelity is lower due to $\langle 2|(a + a_1)^\dagger(a + a_1)|2\rangle > \langle 2|a^\dagger a|2\rangle$.

Finally, let us discuss the optimality of the RL code compared to codewords that are shifted in Fock space, i.e., have the general form $|m\rangle$, $|m + 2\rangle$ (for example, $|1\rangle$, $|3\rangle$ or $|8\rangle$, $|10\rangle$). While the code space better satisfies the KL condition with increasing m, the mean probability of single-photon jumps also increases,

$$\bar{P}_{\text{er}} \propto \frac{1}{4\pi}\int_\Omega \langle\psi_{\theta\phi}|\gamma_a a^\dagger a|\psi_{\theta\phi}\rangle \, \mathrm{d}\Omega = (m + 1)\gamma_a. \tag{20}$$

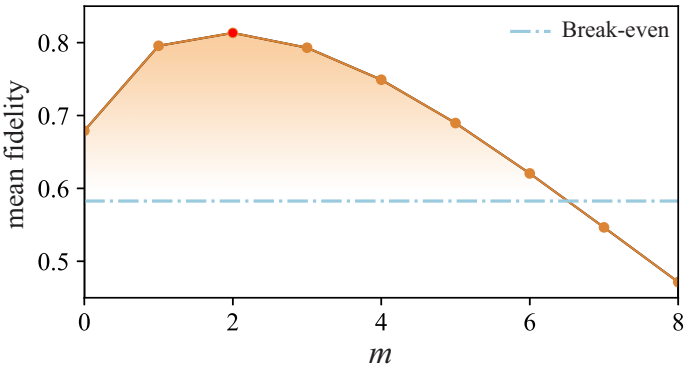

FIG. S5. Simulating the mean fidelity at $t = 150$ μs for different m with the master equation (5) of the main text. Other parameters are $\gamma_a = 0.02$ MHz, $g = 8$ MHz, and $\gamma_b/g = 2.5$.

Therefore, high excitation numbers require large g/γ_a for the jump probability of L_{eng} to be sufficiently large, which increases the difficulty of the experiment. Here we limit the range of the cooperativity $C \leq 160$ to ensure experiment-friendly AQEC. Under this parameter condition, we simulate the mean fidelity for the code word $|m\rangle$, $|m+2\rangle$ in Fig. S5. We find that the mean fidelity of the RL code is optimal compared to its shifted versions. It is worth noting that $m = 0$ is an invalid code space due to $F(\frac{\pi}{2}, \phi, t) < \bar{F}_{\text{be}}$.

S7. SETTING THE COUPLING BETWEEN THE QUBIT & THE ENCODING MODE PLUS THE AUXILIARY MODE

Here we discuss in more detail how to realize our approximate AQEC scheme in an actual physical system. We consider a quantum system consisting of an encoding mode, an auxiliary qubit, and an auxiliary mode. The system Hamiltonian is given by

$$H = \omega_a a^\dagger a + \frac{\omega_b}{2}\sigma_z + \omega_c c^\dagger c + \frac{\chi}{2} a^\dagger a \sigma_z + f(t)(a + a^\dagger)\sigma_x + g_c(t)(c^\dagger + c)\sigma_x, \tag{21}$$

where χ is the dissipative coupling coefficient, $g(t)$ and $f(t)$ are two time-dependent control fields, and ω_a, ω_b, ω_c are the resonant frequencies of the encoding mode, the qubit, and the auxiliary mode, respectively. The system dynamics are then described by the master equation

$$\frac{d\rho}{dt} = -i[H, \rho] + \frac{\gamma_{a1}}{2}\mathcal{D}[a] + \frac{\gamma_{b1}}{2}\mathcal{D}[\sigma_-] + \frac{\gamma_{c1}}{2}\mathcal{D}[c], \tag{22}$$

where γ_{i1} $(i = a, b, c)$ are the corresponding decay rates. The Hamiltonian (21) can be expanded with the eigenstates $|E_{N,M,\pm}\rangle$ of the Hamiltonian $H_0 = \omega_a a^\dagger a + \frac{\omega_b}{2}\sigma_z + \omega_c c^\dagger c + \frac{\chi}{2} a^\dagger a \sigma_z$. Therefore, the Hamiltonian H can be rewritten as

$$\begin{aligned} H = &\sum_{N,M} E_{N,M}|E_{N,M,\pm}\rangle\langle E_{N,M,\pm}| + f(t)\sqrt{N+1}(|E_{N+1,M,+}\rangle\langle E_{N,M,-}| + |E_{N+1,M,-}\rangle\langle E_{N,M,+}| + h.c.) \\ &+ g_c(t)\sqrt{M+1}(|E_{N,M+1,+}\rangle\langle E_{N,M,-}| + |E_{N,M+1,-}\rangle\langle E_{N,M,+}| + h.c.), \end{aligned} \tag{23}$$

where $E_{N,M,\pm} = N\omega_a + M\omega_c \pm (\frac{\omega_b}{2} + \frac{N\chi}{2})$ is the eigenvalue of the Hamiltonian H_0. After switching to a rotating frame by applying the unitary operator $U = \exp(-iH_0 t)$, the Hamiltonian can be written as

$$\begin{aligned} H_I = &\sum_{N,M} f(t)\sqrt{N+1}(|E_{N+1,M,+}\rangle\langle E_{N,M,-}|e^{i(E_{N+1,M,+} - E_{N,M,-})t} + |E_{N+1,M,-}\rangle\langle E_{N,M,+}|e^{i(E_{N+1,M,-} - E_{N,M,+})t}) \\ &+ g_c(t)\sqrt{M+1}(|E_{N,M+1,+}\rangle\langle E_{N,M,-}|e^{i(E_{N,M+1,+} - E_{N,M,-})t} + |E_{N,M+1,-}\rangle\langle E_{N,M,+}|e^{i(E_{N,M+1,-} - E_{N,M,+})t}) + h.c. \,. \end{aligned} \tag{24}$$

Here we assume for simplicity that the mode c and the qubit are resonant, $\omega_b = \omega_c = \omega$, and that the frequencies satisfy $|\omega_a + \omega|, |\omega_a - \omega| \gg \chi \gg g_c(t), |f(t)|$. We can then use the control fields

$$\begin{aligned} f(t) &= \frac{2\alpha_0}{\sqrt{2}}\cos\left[(E_{2,M,+} - E_{1,M,-})\,t\right] + \frac{2\alpha_0}{\sqrt{4}}\cos\left[(E_{4,M,+} - E_{3,M,-})\,t\right], \\ g(t) &= 2\alpha_1\cos\left[(E_{2,M,+} - E_{2,M+1,-})t\right] + 2\alpha_1\cos\left[(E_{4,M,+} - E_{4,M+1,-})t\right], \end{aligned} \tag{25}$$

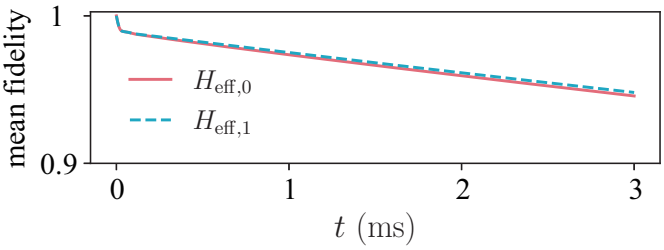

FIG. S6. Evolution of the mean fidelity up to $t = 3$ ms for the effective Hamiltonians $H_{\text{eff},0}$ and $H_{\text{eff},1}$, with the parameters $\alpha_0/2\pi = 0.05$ MHz, $\alpha_1/2\pi = 0.07$ MHz, $\gamma_{a1}/2\pi = 0.2$ kHz, $\gamma_{b1}/2\pi = 2$ kHz, and $\gamma_{c1}/2\pi = 0.12$ MHz.

to select effective transitions. The control fields must satisfy the condition $|f(t)| \leq |g(t)|$ to rapidly transfer the energy from the auxiliary qubit to the auxiliary mode. After performing a rotating-wave approximation, we obtain the following effective Hamiltonian

$$H_{\text{eff},0} \approx \alpha_0(L_o\sigma_+ + L_o^\dagger\sigma_-) + \sum_{N=2,4} \alpha_1|N\rangle\langle N|(c^\dagger\sigma_- + c\sigma_+), \tag{26}$$

which can be further approximated to obtain the simpler expression

$$H_{\text{eff},1} \approx \alpha_0(L_o\sigma_+ + L_o^\dagger\sigma_-) + \alpha_1(c^\dagger\sigma_- + c\sigma_+), \tag{27}$$

due to the fast energy exchange mainly appearing in the code space rather than the error space for the approximate AQEC.

We numerically evaluate the mean fidelity under the two Hamiltonians $H_{\text{eff},0}$ and $H_{\text{eff},1}$ in Fig. (S6), which demonstrates that the above approximation is appropriate. The frame rotation does not affect the decay $D[\sigma_-]$ of the auxiliary mode. Although $D[a]$ and $D[\sigma_-]$ are modified in the rotating frame, the effects of $D[a]$ and $D[\sigma_-]$ remain approximately the same as in the original frame. Finally, we obtain the effective Hamiltonian

$$H_{\text{eff}} \propto L_{\text{eng}}\sigma_+ + L_{\text{eng}}^\dagger\sigma_-, \tag{28}$$

(i.e., the QEC Hamiltonian of the main text) by adiabatically eliminating the high-decay mode c [10]. Therefore, the dynamics described by Eq.(22) are effectively equivalent to the master Eq. (5) of the main text.

[1] Z. Wang, T. Rajabzadeh, N. Lee, and A. H. Safavi-Naeini, PRX Quantum **3**, 020302 (2022).
[2] J. R. Johansson, P. D. Nation, and F. Nori, Comput. Phys. Commun. **183**, 1760 (2012).
[3] J. R. Johansson, P. D. Nation, and F. Nori, Comput. Phys. Commun. **184**, 1234 (2013).
[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, arXiv 10.48550/arXiv.1707.06347 (2017), 1707.06347.
[5] RLlib: Industry-Grade Reinforcement Learning — Ray 3.0.0.dev0 (2022), [Online; accessed 19. Jun. 2022].
[6] P. Zanardi, J. Marshall, and L. C. V., Phys. Rev. A **93**, 022312 (2016).
[7] J.-M. Lihm, K. Noh, and U. R. Fischer, Phys. Rev. A **98**, 012317 (2018).
[8] G. Sarma and H. Mabuchi, New J. Phys. **15**, 035014 (2013).
[9] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).
[10] Y.-C. Liu, X. Luan, H.-K. Li, Q. Gong, C. W. Wong, and Y.-F. Xiao, Phys. Rev. Lett. **112**, 213602 (2014).