# Hybrid Entanglement Distribution between Remote Microwave Quantum Computers Empowered by Machine Learning

Bingzhi Zhang,[1,2,3,†] Jing Wu,[3,4,†] Linran Fan,[4] and Quntao Zhuang[1,3,4,*]

[1] *Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*

[2] *Department of Physics, University of Arizona, Tucson, Arizona 85721, USA*

[3] *Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721, USA*

[4] *James C. Wyant College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA*

(Received 29 June 2022; revised 30 August 2022; accepted 31 October 2022; published 6 December 2022)

Superconducting microwave circuits with Josephson junctions, the major platform for quantum computing, can only reach the full capability when connected. This requires an efficient protocol to distribute microwave entanglement. While quantum computers typically use discrete-variable (DV) methods for information encoding, the entire continuous-variable (CV) degree of freedom in electromagnetic fields must be utilized to achieve the highest entanglement distribution rate. Here, we propose a hybrid protocol to resolve the incompatibility between DV microwave quantum computers and CV quantum communications. CV microwave entanglement is distributed using optical swapping of optical-microwave entanglement pairs. To interface with DV microwave quantum computers, we further design a hybrid circuit to simultaneously convert and distill high-quality DV entanglement from noisy CV entanglement. The hybrid circuit is trained with machine-learning algorithms, ensuring high entanglement fidelity and generation rate. Our work not only provides a practical method to realize efficient quantum links for superconducting microwave quantum computers, but also opens avenues to bridge the gap between DV and CV quantum systems.

DOI: 10.1103/PhysRevApplied.18.064016

## I. INTRODUCTION

Empowered by the law of quantum mechanics, quantum computers have the potential of speeding up the solution of various classically hard problems [1–4]. Among the candidate platforms for quantum computing, superconducting circuits with Josephson junctions stand out with high scalability and strong single-photon nonlinearity, as exemplified by recent demonstrations [5–8]. To unleash their full capabilities, quantum computers need to be connected via entanglement [9–11].

Although direct microwave links can be used at short distances for proof-of-principle demonstrations [12,13], the connection between superconducting quantum computing circuits over long distances is best implemented with optical photons via microwave-optical transduction [14–22]. Only optical photons can maintain the quantum coherence with the low dissipation and decoherence rate at room temperature. In spite of significant improvement, it is still challenging to realize high-performance microwave-optical transduction with near-unity efficiency and near-zero noise. Then a critical question is—what is the most efficient way to connect superconducting microwave quantum computers given the nonideal performance of microwave-optical transduction.

To answer this question, a fundamental issue to be resolved is the gap between continuous-variable (CV) and discrete-variable (DV) quantum information processing: the theory of quantum channel capacity indicates that the entire CV degree of freedom must be utilized to achieve a high entanglement rate [23]; on the other hand, microwave quantum computers typically use a DV approach for information encoding. Moreover, entanglement distillation protocols that improve entanglement quality are also designed separately for DV encoding [24–32] or CV encoding [33–43].

In this paper, we propose a hybrid CV-DV protocol to connect superconducting microwave quantum computers. CV microwave entanglement is first established through optical swapping of optical-microwave entanglement pairs. Then, we develop two entanglement conversion and distillation protocols to transform the noisy CV entanglement to high-fidelity DV entanglement, which can directly interface with transmon qubits. The first protocol adopts the direct CV-to-DV conversion followed by

*qzhuang@usc.edu

†These two authors contributed equally.

 

traditional DV entanglement distillation. The second protocol is based on a machine-learning-enabled hybrid local variational quantum circuit (LVQC) to complete the conversion and distillation simultaneously, which adds to the general toolbox of entanglement manipulation. Estimated with practical nonideal device parameters, both protocols show a huge rate advantage compared with single-photon-based pure DV protocols. The LVQC protocol further shows more than tenfold improvement for the fidelity-success-probability trade-off, compared with the direct CV-to-DV conversion protocol.

## II. SYSTEM SETUP

The overall interconnect protocol consists of two steps, a CV entanglement generation step and a CV-to-DV hybrid conversion and distillation step, as shown in Fig. 1(a). The ultimate goal is to generate high-fidelity entanglement at a high rate between two superconducting microwave quantum computers. To begin with, entanglement is generated between the transduction microwave and optical ports ($\{\hat{a}_1, \hat{m}_1\}$ and $\{\hat{a}_2, \hat{m}_2\}$) on both sides. The optical modes $\hat{a}_1$ and $\hat{a}_2$ then travel through optical links to a center node for entanglement swap to generate noisy CV entanglement between microwave modes $\hat{m}_1$ and $\hat{m}_2$. Next, each microwave mode is coupled into a conversion and distillation module implemented with superconducting quantum computing circuits. Then high-fidelity DV entanglement can be generated from two remote microwave modes. While our protocol is general, we consider the special case of transmon qubits for the wide use in superconducting quantum computing. The LVQC conversion and distillation step is based on the controlled operation on the microwave modes using transmon qubits [Fig. 1(b)], and the final DV entanglement is also between two transmon qubits ($\{q_1, q_2\}$).

For simplicity, we assume identical configuration and performance for the two quantum computing and transduction systems. We also want to point out that we are considering a system with tunable mode-qubit coupling so that the entanglement generation step and entanglement conversion and distillation step can be considered separately. At the same time, we are assuming steady-state operation with perfect spectral-temporal mode matching.

## III. ENTANGLING MICROWAVE MODES VIA CV SWAP

While our protocol applies to general transduction systems, we focus on cavity electro-optic systems [17,18,44,45] for its simplicity as it does not involve intermediate excitations. A typical cavity electro-optic system is shown in Fig. 1(a), where the optical cavity with $\chi^{(2)}$ nonlinearity is placed between the capacitors of a $LC$ microwave resonator. The electric field of the microwave mode modulates the optical resonant frequency across the capacitor via changing the refractive index of the optical cavity. Due to the mixing (rectification) between optical pump and signal in $\chi^{(2)}$ material, microwave field can be generated. The interaction Hamiltonian has the standard three-wave mixing form between two optical modes ($\hat{a}_\ell$ and $\hat{b}_\ell$) and one microwave mode ($\hat{m}_\ell$) [17,18,23,44,45],

$$\hat{H}_\ell = i\hbar(g\hat{a}_\ell^\dagger \hat{b}_\ell \hat{m}_\ell^\dagger - g^* \hat{a}_\ell \hat{b}_\ell^\dagger \hat{m}_\ell), \tag{1}$$

with $g$ the coupling coefficient and $\ell = 1, 2$ for each side. When the optical mode $\hat{b}_\ell$ is coherently pumped, the optical mode $\hat{a}_\ell$ and the microwave mode $\hat{m}_\ell$ will be entangled in a noisy two-mode squeezed vacuum (TMSV) state $\hat{\rho}_{m,o}$, with zero mean and a covariance matrix [23]

$$V_{m,o} = \frac{1}{2}\begin{pmatrix} u\mathbf{I}_2 & v\mathbf{Z}_2 \\ v\mathbf{Z}_2 & w\mathbf{I}_2 \end{pmatrix}, \tag{2}$$

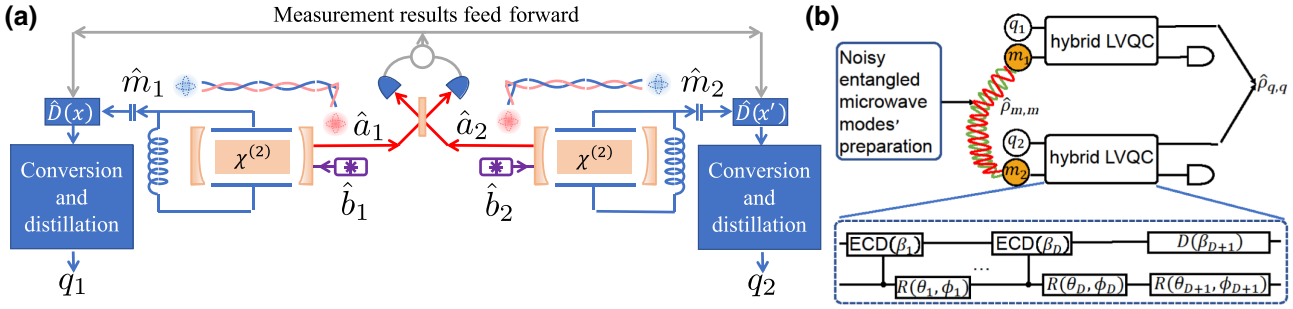


FIG. 1. (a) Interconnect system between two microwave quantum computers. Two cavities enable the generation of microwave-optical entanglement. The optical modes are detected for entanglement-swap, generating microwave-microwave entanglement after displacement $\hat{D}$. Finally, the transmon qubits interact with the microwave modes to generate Bell pairs. (b) Schematic of hybrid local variational quantum circuits (hybrid LVQCs) to distill entanglement from noisy entangled microwave modes $m_1, m_2$ to transmon qubits $q_1, q_2$. The hybrid LVQC is shown in detail in the dashed box with $D$ echoed conditional displacement (ECD) blocks (ECD gates and single qubit rotations) followed by displacement and rotation in the end.

where $\mathbf{Z}_2, \mathbf{I}_2$ are Pauli-$Z$ and identity matrix, and

$$u = 1 + \frac{8\zeta_m[C + n_{\text{in}}(1-\zeta_m)]}{(1-C)^2}, \tag{3a}$$

$$v = \frac{4\sqrt{\zeta_o\zeta_m C}[1 + C + 2n_{\text{in}}(1-\zeta_m)]}{(1-C)^2}, \tag{3b}$$

$$w = 1 + \frac{8C\zeta_o\,[1 + n_{\text{in}}\,(1-\zeta_m)]}{(1-C)^2}. \tag{3c}$$

Here $\zeta_o$ and $\zeta_m$ are the extraction efficiencies for the optical and microwave modes, respectively. The cooperativity $C \propto g^2$ describes the interaction strength [23]. The optical thermal noise is neglected due to its small occupation while the microwave thermal noise has nonzero mean occupation number $n_{\text{in}}$. We also note that any additional optical transmission loss $1-\eta$ can be absorbed into $\zeta_o$ by replacing $\zeta_o$ with $\eta\zeta_o$ (see Appendix D).

To generate entanglement between remote microwave modes, we adopt an entanglement-swap approach. As shown in Fig. 1(a), consider two pairs of entangled microwave-optical modes $\{\hat{a}_1, \hat{m}_1\}$ and $\{\hat{a}_2, \hat{m}_2\}$, each with the covariance matrix Eq. (2). One can interfere the optical modes $\hat{a}_1$ and $\hat{a}_2$ on a balanced beam splitter and perform homodyne detection on the output. Conditioned on the homodyne results, displacement operations will be applied to the two microwave modes, respectively. After the optical homodyne detection and conditioned displacement operation, the microwave modes $\hat{m}_1$ and $\hat{m}_2$ form a noisy TMSV state $\hat{\rho}_{m,m}$ with the covariance matrix

$$V_{m,m} = \frac{1}{2}\begin{pmatrix} \left(u - \frac{v^2}{2w}\right)\mathbf{I}_2 & \frac{v^2}{2w}\mathbf{Z}_2 \\ \frac{v^2}{2w}\mathbf{Z}_2 & \left(u - \frac{v^2}{2w}\right)\mathbf{I}_2 \end{pmatrix}. \tag{4}$$

In the ideal lossless case with $\zeta_o = \zeta_m = 1$, the state $\hat{\rho}_{m,m}$ becomes a pure TMSV state.

We first derive the ultimate rate of entanglement generation, as all later steps are local operations and classical communications (LOCCs), which does not increase the entanglement rate [46]. To characterize the distillable entanglement, we calculate the upper bound by entanglement of formation (EOF) [24] and the lower bound by reverse coherent information (RCI) [47]. The unit of the rate is ebit per round, and can be understood as the following: a rate of $r$ ebit per round means that in $M \gg 1$ rounds, asymptotically $Mr$ pairs of Bell states can be obtained if arbitrary encoding and decoding is allowed.

To begin with, we consider the rate versus the cooperativity in Fig. 2. In the ideal case $\zeta_o = \zeta_m = 1$, EOF and RCI are equal and reduced to entanglement entropy [green line in Fig. 2(a)]. For comparison, we also evaluate the EOF and RCI rate bounds of a pure DV protocol based on the time-bin entanglement [48], where microwave-optical single-photon entanglement is generated by postselecting

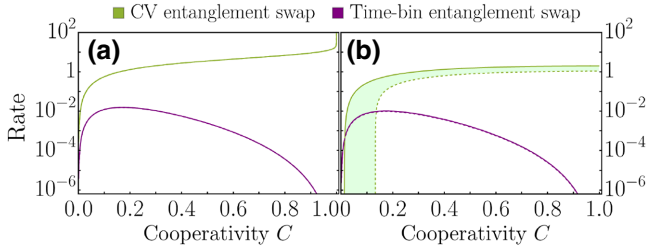


FIG. 2. Rate (in ebit per round) comparison (a) $\zeta_m = \zeta_o = 1$. (b) $\zeta_m = 0.95$, $\zeta_o = 0.9$, and $n_{\text{in}} = 0.2$. The rate of CV entanglement swap is within the green shaded region and the rate of time-bin entanglement swap is within the purple region.

on the state $\hat{\rho}_{m,o}$ (see Appendix C). For all cooperativity values, the proposed CV scheme has more than a 2 orders of magnitude rate advantage over the pure DV protocol based on time-bin entanglement. In Fig. 2(b), we further consider the nonideal extraction with $\zeta_m = 0.95$ and $\zeta_o = 0.9$, and nonzero microwave thermal noise $n_{\text{in}} = 0.2$ (corresponding to approximately 0.2 Kelvin temperature at 8 GHz). Although rigorous advantage happens only at cooperativity above 0.12, we expect the lower bound to be nontight and actual advantage should still be large in the low cooperativity region. In Appendix D, we show that such a rate advantage can be identified for smaller $\zeta_o$, and therefore is robust against optical transmission loss. We note a recent paper [49] showing similar rate advantage with a different platform and approach.

## IV. CONVERSION AND DISTILLATION PROTOCOLS

Next, we design protocols to obtain high-fidelity DV entanglement $\hat{\rho}_{q,q}$ between two transmon qubits from the noisy two-mode squeezing between microwave modes $\hat{\rho}_{m,m}$. Distillation has been explored separately for either DV [24–32] or CV [33–43] entanglement. Here, we require a hybrid conversion and distillation protocol that produces high-fidelity DV entanglement from noisy CV entanglement. We present two different approaches: (i) direct swap where conversion and distillation are separate; and (ii) hybrid LVQC where universal control completes both conversion and distillation simultaneously. With symmetric configuration between the two ends of the interconnect system, we omit subscripts in the following. As practical entanglement distillation is by nature probabilistic, the performance is quantified by two metrics—the success probability and the fidelity conditioned on success.

### A. Direct conversion

We begin with a simple direct conversion from CV to noisy entangled qubit pairs. Inspired by Ref. [50], we consider the interaction between the microwave mode and a

transmon qubit with Hamiltonian

$$\hat{H}_{\text{swap}} = \hat{m} \otimes |1\rangle\langle 0|_q + \text{h.c.}, \tag{5}$$

We can control the interaction time $t$ such that the unitary operator $\hat{U} = \exp(-it\hat{H}_{\text{swap}})$ gives the maximum EOF between the qubits from the two sides after disregarding the corresponding microwave modes (see Appendix E). This will produce a noisy entangled qubit pair deterministically. One can then follow up with further DV distillation protocols. Later we show that such a simple scheme is able to convert most of the entanglement already, therefore we do not consider more complicated CV-to-DV conversion schemes [51].

### B. Variational hybrid conversion and distillation

To explore the ultimate performance, we further use a variational approach to design a hybrid conversion and distillation protocol, as shown in Fig. 1(b). Two qubits are initialized in $|0\rangle_q$, and put into interaction with the two microwave modes, respectively. The hybrid LVQC includes a series of single-qubit rotation and echoed conditional displacement (ECD), followed by another displacement at the end [8]. Each single-qubit rotation is characterized by two angles $\phi$ and $\theta$,

$$\hat{R}(\theta,\phi) = \exp[-i(\theta/2)(\hat{\sigma}_x \cos\phi + \hat{\sigma}_y \sin\phi)], \tag{6}$$

where $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are the Pauli operators. Each ECD gate acts on the microwave mode $\hat{m}$ and the qubit $\hat{q}$ as

$$\hat{U}_{\text{ECD}}(\beta) = \hat{D}\left(\beta/2\right) \otimes |1\rangle\langle 0|_q + \text{h.c.}, \tag{7}$$

where $\hat{D}(\alpha) \equiv \exp(\alpha\hat{m}^\dagger - \alpha^*\hat{m})$ is the displacement operator. Compared with previous proposals for universal control [52,53], the considered ECD gate approach has advantage in both control speed and fidelity [8].

To determine the success of entanglement conversion and distillation, we perform measurements characterized by positive operator-valued measure (POVM) $\{\hat{\Pi}_s, \hat{\mathbf{I}} - \hat{\Pi}_s\}$. The success probability is therefore

$$P_{\text{success}} = \text{Tr}\left\{\hat{\Pi}_s \left(\hat{U}_D \otimes \hat{U}_D\right) \hat{\rho}_{qm,qm} \left(\hat{U}_D^\dagger \otimes \hat{U}_D^\dagger\right)\right\}, \tag{8}$$

where $\hat{\rho}_{qm,qm}$ is the composite initial quantum state including both microwave modes and qubits on the two sides and $\hat{U}_D$ is the unitary of the LVQC with $D$ layers of ECD and single-qubit rotations. We choose photon counting on the microwave modes and postselect on photon number lower than certain threshold $n_{\text{th}}$, so $\Pi_s = \sum_{i,j=0}^{n_{\text{th}}-1} |i\rangle\langle i|_{m_1} \otimes |j\rangle\langle j|_{m_2}$, where $|k\rangle$ is the number state. In this paper, we choose $n_{\text{th}} = 5$, which is half of the photon number cut off in simulation. Due to the variational training, we expect the performance will not depend the photon-number threshold, as long as it is not too small. The state of transmon qubits conditioned on successful conversion and distillation is given by

$$\hat{\rho}_{q,q} \propto \text{Tr}_{m,m}\left\{\hat{\Pi}_s \left(\hat{U}_D \otimes \hat{U}_D\right) \hat{\rho}_{qm,qm} \left(\hat{U}_D^\dagger \otimes \hat{U}_D^\dagger\right) \hat{\Pi}_s\right\}. \tag{9}$$

We train the hybrid LVQC towards distilling a perfect Bell pair $|\Psi^+\rangle \equiv (|0\rangle_q |0\rangle_q + |1\rangle_q |1\rangle_q)/\sqrt{2}$. The cost function is defined with the success probability $P_{\text{success}}$ and fidelity $F\left(\hat{\rho}_{qq}\right) = \langle\Psi^+|\,\hat{\rho}_{q,q}\,|\Psi^+\rangle$, as

$$\mathcal{C}\left(\{\beta\},\{\theta\},\{\phi\}\right) = (1 - P_{\text{success}}) + \lambda \times \text{Softplus}(F_c - F), \tag{10}$$

where the penalty coefficient $\lambda$ and critical fidelity $F_c$ are hyperparameters to tune the trade-off between success probability and fidelity. The softplus function $\text{Softplus}(x) \equiv \log\left(1 + e^{\gamma x}\right)/\gamma$ with $\gamma = 20$ is introduced to enable a smooth penalty on the $F \leq F_c$ events.

Both the direct swap and hybrid LVQC can be operated on a single copy of $\hat{\rho}_{m,m}$ to generate a single copy ($M = 1$) of entangled qubit pair. To further improve their performance, we also consider a two-copy case ($M = 2$) where two copies of noisy CV entanglement $\hat{\rho}_{m,m}$ are first converted to two pairs of noisy entangled qubits, and then further DV distillation is performed on the two noisy qubit pairs to produce the final entangled state. In the last step, we consider both the traditional Deutsch-Ekert-Jozsa-Macchiavello-Popescu-Sanpera (DEJMPS) protocol [26] and a DV LVQC protocol. The DV LVQC protocol utilizes a standard universal gate set of single-qubit rotations and CNOTs, with Pauli-$Z$ measurement providing the postselection and the same cost function as Eq. (10), as we detail in Appendix F.

### C. Performance comparison

We begin our performance comparison with the trade-off between infidelity $1 - F$ and the success probability $P_{\text{success}}$. As shown in Fig. 3(a), the simple direct swap produces $1 - F \sim 0.23$ deterministically (red cross). The performance of any protocol following the direct swap will be bounded by the positive-partial-transpose (PPT) bound (red dashed), which can be numerically evaluated by a semidefinite program [27]. In contrast, the hybrid LVQC approach directly achieves a 1 order of magnitude advantage in the infidelity (blue solid). Indeed, when applying the PPT bound on $\hat{\rho}_{m,m}$, the (generally loose) lower bound of infidelity for LVQC protocols (black dashed) also decreases substantially, compared with the PPT bound from the direct swap (red dashed).

To further lower the infidelity, we consider the two-copy case, where DV processing is further performed on the

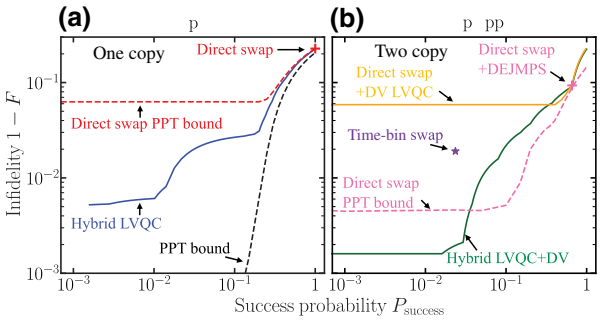


FIG. 3. Infidelity of transmon qubits versus success probability on (a) one-copy and (b) two-copy entangling microwave modes with $C = 0.1, n_{\text{in}} = 0.2, \zeta_m = 0.992, \zeta_o = 0.99$. The PPT bound (black) in (a) goes below $10^{-4}$ and saturates; red and pink dashed curves in (a),(b) correspond to PPT bound of one-copy and two-copy two-qubit state by direct swap. There are $D = 10$ ECD blocks in hybrid LVQCs and $L = 6$ layers in DV LVQCs.

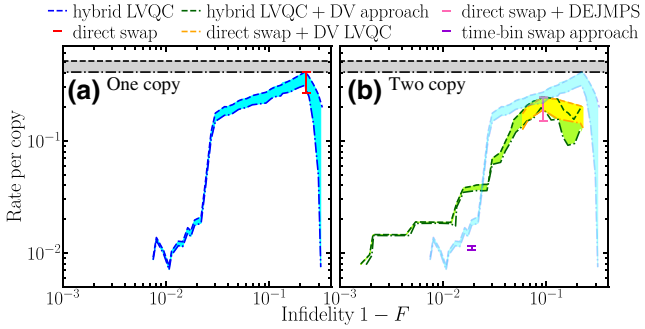


FIG. 4. Entanglement rate per copy on (a) one-copy and (b) two-copy noisy entangled microwave modes versus infidelity $1 - F$, at identical parameters to Fig. 3. Dot-dashed and dashed curves correspondingly represent RCI lower bounds and EOF upper bounds. Shaded areas and line segments show the range of rate. The shallow blue curves and areas in (b) are the same as (a) for comparison.

output of the one-copy case. We first consider the traditional DEJMPS following the direct swap protocol as the reference [pink cross in Fig. 3(b)]. When the traditional DEJMPS is replaced by DV LVQC (orange line), lower infidelity can be achieved with lower success probability. When we use DV LVQC to further distill the two-qubit pairs from the hybrid LVQC, a 2 orders of magnitude lower infidelity (green) can be achieved. Note that even compared with the (generally loose) two-copy PPT lower bound of the direct swap approach (pink dashed), advantages can still be identified in the low infidelity region. As a benchmark, the time-bin-based entanglement swap (purple star, see Appendix C) is 1 order of magnitude worse in terms of the infidelity, when compared with the hybrid LVQC (green) at the same success probability.

Although we consider only two copies, the advantage of the hybrid LVQC scheme over the direct swap scheme can be generalized to more copies. As the hyrid LVQC protocol starts with better entanglement generation performance [Fig. 3(a)], it will also perform better when further distillation steps are implemented.

Finally, we compare the entanglement rate per copy for a $M$-copy conversion and distillation protocol,

$$R_E(\hat{\rho}_{q,q}) = P_{\text{success}} \times E\left(\hat{\rho}_{q,q}\right)/M \leq E\left(\hat{\rho}_{m,m}\right), \qquad (11)$$

where $E(\cdot)$ is the distillable entanglement. As mentioned before, we utilize the lower and upper bounds of distillable entanglement, RCI and EOF for evaluation. In Fig. 4, we see that the rate in general decays as the required infidelity decreases; this is because we are considering a fixed number of copies, which is far from an asymptotic limit. In the one-copy case, the hybrid LVQC approach (blue) is able to achieve close to optimal rate while improving the fidelity up to 0.97, while further improving fidelity drastically decreases the rate; at the same time, direct swap (red) achieves only a single point with a large infidelity and a similar rate, compared with a versatile rate-infidelity trade-off enabled by LVQC. In the two-copy case, hybrid LVQC with additional DV postprocessing is able to improve the infidelity to approximately $10^{-3}$ while the rate is kept approximately $10^{-2}$ (green). At higher infidelity, hybrid LVQC is also as good as the other approaches (orange) with direct swap and additional DV postprocessing (including a DV LVQC circuit). Most notably, at the same level of infidelity, the rate of the hybrid LVQC approach (green) is a factor of 3.4 higher than the time-bin entanglement swap approach (purple dot), as shown in Fig. 4(b).

In Fig. 4(b), we also compare the two-copy results with the one-copy hybrid LVQC results (shallow blue). It is interesting to note that the one-copy protocol has higher rates for the same infidelity when $1 - F \gtrsim 0.024$. This is mainly due to the fact that in a two-copy protocol we need to postselect on two one-copy success events, which substantially reduces the overall success probability. However, if one wants to reach substantially lower infidelity, then two-copy protocols give a much better rate.

In practice, the transmon qubits in the hybrid system can undergo various types of noise and decoherence. To bring an insight to the noisy scenario of hybrid LVQCs, we consider the depolarizing noise accompanied with each qubit operation. Such noise is modeled by a depolarization channel $\Delta_p$ with parameter $p$, which takes an input single-qubit state $\hat{\rho}$ to the output state contaminated by the fully mixed state $\hat{I}/2$ as $\Delta_p(\hat{\rho}) = (1-p)\hat{\rho} + p\hat{I}/2$. In short, for a $D$-layer hybrid LVQC, the qubits undergo $D + 1$ depolarization channels. To understand the performance decay with noise, we consider the same hybrid LVQC setup of Fig. 3 at the operating point of infidelity at $1 - F \sim 0.02$ and success probability $P_{\text{success}} \sim 0.017$. The simulation results are shown in Fig. 5. Starting from

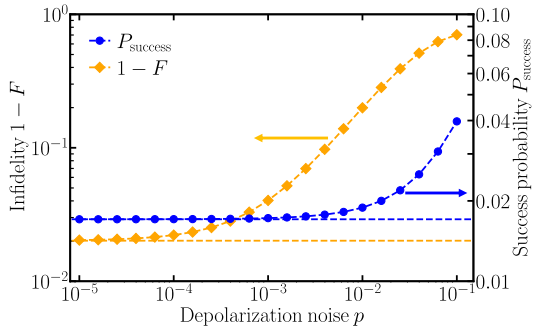


FIG. 5. Infidelity of transmon qubits (orange) and distillation success probability (blue) versus depolarization noise on one-copy entangling microwave modes at identical parameters as in Fig. 3. Orange and blue horizontal lines indicate the infidelity and success probability without noise.

the noiseless circuit achieving infidelity at $1-F\sim 0.02$ (orange dashed) and success probability $P_{\text{success}}\sim 0.017$ (blue dashed), as the noise become stronger, we see the infidelity grows as expected. While for success probability, as the depolarization channel leads to a mixture with the maximally mixed state, entangling modes through the hybrid LVQCs approaches a uniform superposition of displacements determined by noiseless optimization with stronger noise $p$, therefore the success probability $p$ increases given the fixed final measurement, though slowly compared to the infidelity.

## V. DISCUSSIONS

We propose an interconnect system based on CV entanglement swap and hybrid variational entanglement conversion and distillation to entangle microwave superconducting quantum computers. Our approach provides a huge rate and fidelity advantage compared with the time-bin approach. In particular, the hybrid conversion and distillation protocol provides an infidelity-success-probability trade-off, with an order of magnitude advantage over the direct swap approach. In the multicopy scenario, it is an open problem of how to optimize the circuit design to get closer to the PPT lower bound. While we analyze the depolarizing noise in qubit operations, the impact of other nonideal factors in experiments still needs further investigation. For example, the upper and lower bounds are calculated assuming steady-state operation with perfect spectral-temporal mode matching. Nonperfect interference and cavity reflection can happen with spectral-temporal mismatching, leading to the decrease of the entanglement rate. Therefore, the upper bound is still accurate, but the lower bound should be calculated depending on specific device metrics. Although we consider cavity electro-optics as the transduction system, our protocol also applies to other transduction systems and the analyses can be done in the same way based on a modified interaction Hamiltonian Eq. (2).

## ACKNOWLEDGMENTS

Q.Z. and L.F. acknowledge discussion with Filip Rozpeędek, Chi Xiong, Edward H Chen, Jason S Orcutt, John A Smolin, Vikesh Siddhu, and Abram L Falk. B.Z., J.W., and Q.Z. acknowledge support from NSF CAREER Award CCF-2142882, Defense Advanced Research Projects Agency (DARPA) under Young Faculty Award (YFA) Grant No. N660012014029 and National Science Foundation (NSF) Engineering Research Center for Quantum Networks Grant No. 1941583. F.L. acknowledges support from U.S. Department of Energy UT-Battelle/Oak Ridge National Laboratory (4000178321), Office of Naval Research (N00014-19-1-2190) and National Science Foundation (CCF-1907918).

## APPENDIX A: CONTINUOUS-VARIABLE ENTANGLEMENT SWAP

In this paper, we adopt the same convention of quadratures and covariance matrix as Ref. [23]. For a review of these definitions, refer to Appendices of Ref. [23]. Below we give a simple overview. The Wigner function of a Gaussian state $\hat{\rho}$ is

$$W(\mathbf{x})=\frac{1}{(2\pi)^n|\mathbf{V}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\mathbf{V}^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right], \tag{A1}$$

where $\bar{\mathbf{x}}$ and $\mathbf{V}$ are the quadrature mean and covariance matrix of state $\hat{\rho}$:

$$\bar{\mathbf{x}}\equiv \mathrm{Tr}[\hat{\rho}\hat{\mathbf{x}}], \tag{A2}$$

$$V_{ij}\equiv\frac{1}{2}\mathrm{Tr}[\hat{\rho}\{\hat{x}_i-\bar{x}_i,\hat{x}_j-\bar{x}_j\}]. \tag{A3}$$

Here $\hat{\mathbf{x}}\equiv(\hat{q}_1,\hat{p}_1,\hat{q}_2,\hat{p}_2,\ldots)$ is the vector of quadrature operators.

A Gaussian state is mapped to another Gaussian state under a Gaussian unitary $\hat{U}_{\mathbf{S},\mathbf{d}}$ (described by a symplectic matrix $\mathbf{S}$ and a displacement vector $\mathbf{d}$), with the mean and covariance matrix transformed as

$$\bar{\mathbf{x}}\to\mathbf{S}\bar{\mathbf{x}}+\mathbf{d},\mathbf{V}\to\mathbf{S}\mathbf{V}\mathbf{S}^T. \tag{A4}$$

### 1. Electro-optical system

The entanglement between the optical and microwave fields can be generated from a SPDC process by pumping a triple-resonance device [17]. The system with coupling

strength $g$ is modeled by the total Hamiltonian

$$H = \hbar\omega_o \hat{a}^\dagger \hat{a} + \hbar\omega_m \hat{m}^\dagger \hat{m} + \hbar(g\hat{a}^\dagger \hat{m}^\dagger + g^* \hat{a}\hat{m}), \quad \text{(A5)}$$

where $\hat{a}$ ($\hat{m}$) is the annihilation operator for optical (microwave) field with resonance at frequency $\omega_o$ ($\omega_m$). The intracavity pump power and the phase-matching condition are absorbed into the coupling strength $g$.

The output fields can be solved by Heisenberg-Langevin equations in the Fourier domain with the input-output relations [48,54]

$$0 = \mathbf{G}\hat{\mathbf{a}} + \mathbf{K}\hat{\mathbf{a}}_{\text{in}}, \quad \text{(A6)}$$

$$\hat{\mathbf{a}}_{\text{out}} = -\mathbf{K}^T\hat{\mathbf{a}} + \hat{\mathbf{a}}_{\text{in}}. \quad \text{(A7)}$$

We present the dynamics in the Fourier domain with optical frequency detuning $\Delta_p = \omega - \omega_o$ and microwave frequency detuning $\Delta_e = \omega - \omega_m$. The notations are defined as the following:

$$\hat{\mathbf{a}} = (\hat{a}, \hat{a}^\dagger, \hat{m}, \hat{m}^\dagger)^T, \quad \text{(A8)}$$

$$\hat{\mathbf{a}}_{\text{in}} = (\hat{a}_{\text{in}}, \hat{a}_{\text{in}}^\dagger, \hat{a}^{(i)}, \hat{a}^{\dagger(i)}, \hat{m}_{\text{in}}, \hat{m}_{\text{in}}^\dagger, \hat{m}^{(i)}, \hat{m}^{\dagger(i)})^T, \quad \text{(A9)}$$

$$\mathbf{G} = \begin{pmatrix} -\frac{\gamma_p}{2} + i\Delta_p & 0 & 0 & -ig \\ 0 & -\frac{\gamma_p}{2} - i\Delta_p & ig & 0 \\ 0 & -ig & -\frac{\gamma_e}{2} + i\Delta_e & 0 \\ ig & 0 & 0 & -\frac{\gamma_e}{2} - i\Delta_e \end{pmatrix}, \quad \text{(A10)}$$

$$\mathbf{K} = \begin{pmatrix} \sqrt{\gamma_{pc}} & 0 & \sqrt{\gamma_{pi}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\gamma_{pc}} & 0 & \sqrt{\gamma_{pi}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{\gamma_{ec}} & 0 & \sqrt{\gamma_{ei}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{\gamma_{ec}} & 0 & \sqrt{\gamma_{ei}} \end{pmatrix}. \quad \text{(A11)}$$

The output fields then relate to the input fields by

$$\hat{\mathbf{a}}_{\text{out}} = \mathbf{S}_a\hat{\mathbf{a}}_{\text{in}} = (\mathbf{K}^T\mathbf{G}^{-1}\mathbf{K} + \mathbf{I}_8)\hat{\mathbf{a}}_{\text{in}}. \quad \text{(A12)}$$

The quadrature observables and the annihilation operators are related by

$$\begin{pmatrix} \hat{q}^a \\ \hat{p}^a \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^\dagger \end{pmatrix}, \quad \text{(A13)}$$

$$\mathbf{Q} = \mathbf{I}_4 \otimes \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \quad \text{(A14)}$$

we get the input-output quadrature relation

$$\hat{\mathbf{x}}_{\text{out}} = \mathbf{S}_x\hat{\mathbf{x}}_{\text{in}} = \mathbf{Q}\mathbf{S}_a\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\text{in}}, \quad \text{(A15)}$$

$$\hat{\mathbf{x}}_{\text{in}} = (\hat{q}^p_{\text{in}}, \hat{p}^p_{\text{in}}, \hat{q}^{p,(i)}, \hat{p}^{p,(i)}, \hat{q}^e_{\text{in}}, \hat{p}^e_{\text{in}}, \hat{q}^{e,(i)}, \hat{p}^{e,(i)})^T. \quad \text{(A16)}$$

Then, the input-output relation of the covariance matrix is derived as

$$\mathbf{V}_{\text{out}} = \mathbf{S}_x\mathbf{V}_{\text{in}}\mathbf{S}_x^T, \quad \text{(A17)}$$

in which the input covariance matrix $\mathbf{V}_{\text{in}}$ contains vacuum noise from the optical modes and the input microwave mode while the dissipation microwave mode is contaminated by the thermal noise of population $n_{\text{in}}$,

$$\mathbf{V}_{\text{in}} = \text{Diag}(\mathbf{I}_6, (n_{\text{in}} + 1/2)\mathbf{I}_2). \quad \text{(A18)}$$

The covariance matrix of two output fields finally shows in the form as

$$V_{o,m} = \frac{1}{2} \begin{pmatrix} w & 0 & 0 & -v \\ 0 & w & -v & 0 \\ 0 & -v & u & 0 \\ -v & 0 & 0 & u \end{pmatrix}, \quad \text{(A19)}$$

where the parameters in Eq. (A19) are as follows:

$$u = 1 + \frac{8\zeta_m[C + n_{\text{in}}(1 - \zeta_m)]}{(1 - C)^2}, \quad \text{(A20a)}$$

$$v = \frac{4\sqrt{\zeta_o\zeta_m C}[1 + C + 2n_{\text{in}}(1 - \zeta_m)]}{(1 - C)^2}, \quad \text{(A20b)}$$

$$w = 1 + \frac{8C\zeta_o\,[1 + n_{\text{in}}\,(1 - \zeta_m)]}{(1 - C)^2}, \quad \text{(A20c)}$$

where the cooperativity is defined as $C = 4g^2/(\gamma_e\gamma_p)$ and extraction coefficients are defined as $\zeta_o = \gamma_{pc}/\gamma_p$ and $\zeta_m = \gamma_{ec}/\gamma_e$ with $\gamma_e = \gamma_{ec} + \gamma_{ei}$ and $\gamma_p = \gamma_{pc} + \gamma_{pi}$. Rotating the

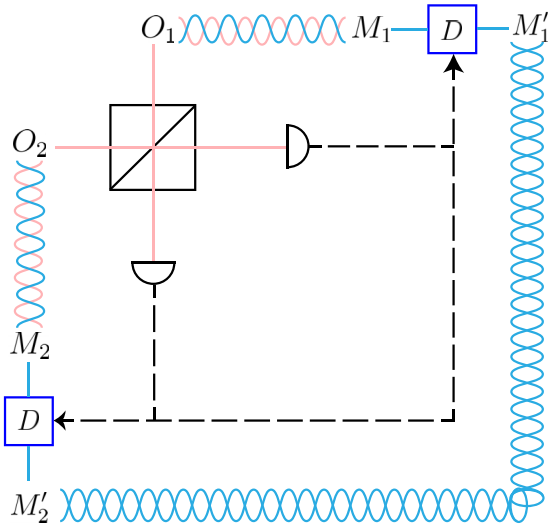


FIG. 6. Continuous-variable entanglement swap scheme. The cavities generate two entangled mode pairs $O_1M_1$ and $O_2M_2$. Then two optical modes $O_1$ and $O_2$ are mixed by a balanced beam splitter. We perform homodyne measurements at the output of the beam splitter and then displace the microwave modes $M_1$ and $M_2$, producing two entangled microwave modes $M_1'$ and $M_2'$.

microwave mode by a constant phase $\pi/2$ we get the covariance matrix in a standard form

$$V_{MO} = \frac{1}{2}\begin{pmatrix} u\mathbf{I}_2 & v\mathbf{Z}_2 \\ v\mathbf{Z}_2 & w\mathbf{I}_2 \end{pmatrix}. \tag{A21}$$

### 2. Entanglement swap

The CV entanglement scheme is shown in Fig. 6. Each of the entangled mode pairs $O_1M_1$ (with annihilation operators $\{\hat{a}_1, \hat{m}_1\}$) and $O_2M_2$ (with annihilation operators $\{\hat{a}_2, \hat{m}_2\}$) from the cavities are in a noisy TMSV state $\hat{\rho}_{m,o}$, with zero mean and the covariance matrix $V_{m,o}$ given by Eq. (2) in the main text.

The composite four-mode system $M_1M_2O_1O_2$ in the entanglement swap starts in a product of two identical Gaussian states $\hat{\rho}_{m,o}$ with an overall covariance matrix

$$V_{M_1M_2O_1O_2} = \frac{1}{2}\begin{pmatrix} u\mathbf{I}_2 & \mathbf{0} & v\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & u\mathbf{I}_2 & \mathbf{0} & v\mathbf{Z}_2 \\ v\mathbf{Z}_2 & \mathbf{0} & w\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & v\mathbf{Z}_2 & \mathbf{0} & w\mathbf{I}_2 \end{pmatrix}. \tag{A22}$$

First, one interferes the optical modes $O_1$ (with annihilation operator $\hat{a}_1$) and $O_2$ (with annihilation operator $\hat{a}_2$) with a balanced beam splitter, described by the symplectic transform

$$\mathbf{S}_{\rm BS} = \mathbf{I}_4 \otimes \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{I}_2 & \mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 \end{pmatrix}. \tag{A23}$$

The covariance matrix after the beam splitter

$$V'_{M_1M_2O_1O_2} = \mathbf{S}_{\rm BS} V_{M_1M_2O_1O_2} \mathbf{S}_{\rm BS}^T \equiv \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_3 \\ \mathbf{C}_3^T & \mathbf{C}_2 \end{pmatrix}, \tag{A24}$$

where we denote it in a block form. The position quadrature $\hat{q}_-$ and the momentum quadrature $\hat{p}_+$ of the outputs of the beam splitter are measured by homodyne. Let $\tilde{\mathbf{x}} = \sqrt{2}(q_-, -p_+)^T$ be the rescaled result of homodyne measurements, and $\Pi_4 = \mathrm{Diag}(0,1,1,0)$ be the projection to integrate out the other variables in the Wigner function. The state of $M_1M_2$ after the homodyne measurements is described by its covariance matrix

$$\begin{aligned} V_{M_1M_2} &= \mathbf{C}_1 - \mathbf{C}_3(\Pi_4\mathbf{C}_2\Pi_4)^{-1}\mathbf{C}_3^T, \\ &= \frac{1}{2}\begin{pmatrix} (u - \frac{v^2}{2w})\mathbf{I}_2 & \frac{v^2}{2w}\mathbf{Z}_2 \\ \frac{v^2}{2w}\mathbf{Z}_2 & (u - \frac{v^2}{2w})\mathbf{I}_2 \end{pmatrix}, \end{aligned} \tag{A25}$$

and mean

$$\bar{\mathbf{x}}_{12} = \left(-\frac{v}{2w}\mathbf{Z}_2\tilde{\mathbf{x}}, \frac{v}{2w}\tilde{\mathbf{x}}\right). \tag{A26}$$

When displacements $\hat{D}(v/2w\mathbf{Z}_2\tilde{\mathbf{x}})$ and $\hat{D}(-v/2w\tilde{\mathbf{x}})$ are applied at modes $M_1$ and $M_2$, respectively, the final output state of system $M_1'M_2'$ (with annihilation operators $\hat{m}_1, \hat{m}_2$) is a zero-mean Gaussian state with the same covariance matrix as $V_{M_1M_2}$ in Eq. (A25). (For the definition of the Weyl displacement operator, refer to Ref. [23].)

### 3. Teleportation approach

The other CV-teleportation approach is shown in Fig. 7. In this approach, one starts with a strong TMSV pair $O_3O_4$ in the optical domain, with covariance matrix

$$V_{O_3O_4} = \frac{1}{2}\begin{pmatrix} \cosh(2r)\mathbf{I}_2 & \sinh(2r)\mathbf{Z}_2 \\ \sinh(2r)\mathbf{Z}_2 & \cosh(2r)\mathbf{I}_2 \end{pmatrix}. \tag{A27}$$

The covariance matrix of the composite system $M_1M_2O_1O_3O_2O_4$ is

$$V = \frac{1}{2}\begin{pmatrix} u\mathbf{I}_2 & \mathbf{0} & v\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & u\mathbf{I}_2 & \mathbf{0} & \mathbf{0} & v\mathbf{Z}_2 & \mathbf{0} \\ v\mathbf{Z}_2 & \mathbf{0} & w\mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cosh(2r)\mathbf{I}_2 & \mathbf{0} & \sinh(2r)\mathbf{Z}_2 \\ \mathbf{0} & v\mathbf{Z}_2 & \mathbf{0} & \mathbf{0} & w\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sinh(2r)\mathbf{Z}_2 & \mathbf{0} & \cosh(2r)\mathbf{I}_2 \end{pmatrix}. \tag{A28}$$

Two balanced beam splitters are applied to $O_1O_3$ and $O_2O_4$ separately:

$$\begin{aligned} \mathbf{S}_{\rm BS1} &= \mathbf{I}_2 \otimes \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{I}_2 & \mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 \end{pmatrix} \otimes \mathbf{I}_2, \\ \mathbf{S}_{\rm BS2} &= \mathbf{I}_4 \otimes \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{I}_2 & \mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 \end{pmatrix}, \end{aligned}$$

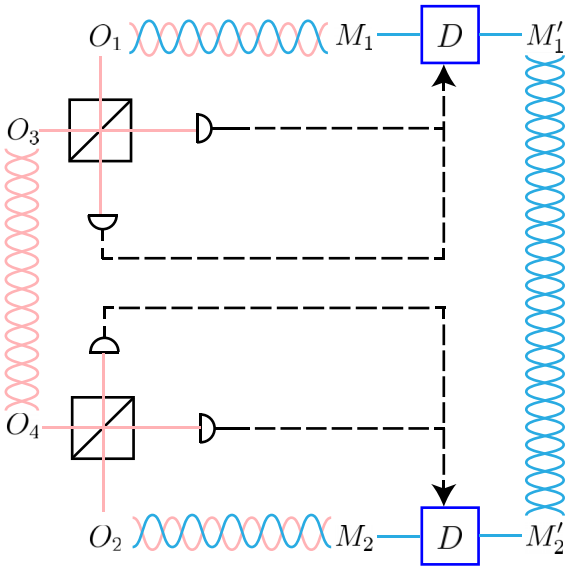


FIG. 7. CV teleportation scheme. We start with a TMSV state between optical mode $O_3$ and $O_4$. Then they will be teleported by use of two $M$-$O$ pairs. After teleportation, the microwave modes $M_1'$ and $M_2'$ are entangled.

leading to the output

$$V' = \mathbf{S}_{\mathrm{BS2}}\mathbf{S}_{\mathrm{BS1}} V \mathbf{S}_{\mathrm{BS1}}^T \mathbf{S}_{\mathrm{BS2}}^T, \equiv \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_3 \\ \mathbf{D}_3^T & \mathbf{D}_2 \end{pmatrix}. \tag{A29}$$

Let $\Pi_8 \equiv \mathrm{Diag}(0,1,1,0,0,1,1,0)$ be the projection, $\tilde{\mathbf{x}}^{(1)} \equiv \sqrt{2}(q_-^{(3)}, -p_+^{(1)})^T$ be measurement results on $O_1O_3$ port and $\tilde{\mathbf{x}}^{(2)} \equiv \sqrt{2}(q_-^{(4)}, -p_+^{(2)})^T$ be measurement results on $O_2O_4$ port. Then the covariance matrix and mean of the output microwave modes $M_1M_2$ are

$$V_{12} = \mathbf{D}_1 - \mathbf{D}_3(\Pi_8\mathbf{D}_2\Pi_8)^{-1}\mathbf{D}_3^T = \frac{1}{2} \times \begin{pmatrix} (u - \frac{v^2[w+\cosh(2r)]}{1+w^2+2w\cosh(2r)})\mathbf{I}_2 & \frac{v^2\sinh(2r)}{1+w^2+2w\cosh(2r)}\mathbf{Z}_2 \\ \frac{v^2\sinh(2r)}{1+w^2+2w\cosh(2r)}\mathbf{Z}_2 & (u - \frac{v^2[w+\cosh(2r)]}{1+w^2+2w\cosh(2r)})\mathbf{I}_2 \end{pmatrix}, \tag{A30}$$

$$\bar{\mathbf{x}}_{12} = (-\mathbf{Z}_2\tilde{\mathbf{x}}, \tilde{\mathbf{x}}), \text{ with } \tilde{\mathbf{x}} = \frac{v\sinh(2r)}{1+w^2+2w\cosh(2r)}\tilde{\mathbf{x}}^{(1)} - \frac{v[w+\cosh(2r)]}{1+w^2+2w\cosh(2r)}\mathbf{Z}_2\tilde{\mathbf{x}}^{(2)}. \tag{A31}$$

We can then perform displacements to cancel the mean and generate the final entangled microwave system $M_1'M_2'$, with zero-mean and covariance matrix given by Eq. (A30). The output of the second scheme reduces to that of the first scheme when $O_3O_4$ are in the infinite squeezing limit ($r \to +\infty$).

## APPENDIX B: ENTANGLEMENT MEASURES

The evaluation of distillable entanglement in general requires a regularization of an infinite number of copies; therefore, as explained in the main text, we consider the lower and upper bounds of it, RCI and EOF [55,56]. For a bipartite quantum system with $A$ and $B$ in a state $\hat{\rho}_{AB}$, if classical communication is allowed, then the rate of entanglement generation is lower bounded by RCI [47]

$$I_R(\hat{\rho}_{AB}) = \max\{S(\hat{\rho}_B), S(\hat{\rho}_A)\} - S(\hat{\rho}_{AB}), \tag{B1}$$

where $S(\hat{\rho}) = -\mathrm{Tr}\hat{\rho}\log_2\hat{\rho}$ is the von Neumman entanglement entropy of state $\hat{\rho}$ and $\hat{\rho}_A = \mathrm{Tr}_B\hat{\rho}_{AB}$ is the reduced density matrix of subsystem $A$ and similarly for $\hat{\rho}_B$.

For a symmetric two-mode Gaussian state characterized by the covariance matrix

$$V = \frac{1}{2}\begin{pmatrix} a\mathbf{I}_2 & c\mathbf{Z}_2 \\ c\mathbf{Z}_2 & b\mathbf{I}_2 \end{pmatrix}, \tag{B2}$$

in the standard form, its symplectic eigenvalues are simply given as $\nu_\pm = [\sqrt{y} \pm (b-a)]/2$ with $y \equiv (a+b)^2 - 4c^2$ [57]. The entanglement entropy is $S(\hat{\rho}) = g(\nu_+) + g(\nu_-)$, where

$$g(x) \equiv \frac{x+1}{2}\log_2\left(\frac{x+1}{2}\right) - \frac{x-1}{2}\log_2\left(\frac{x-1}{2}\right). \tag{B3}$$

For the microwave modes in Eq. (4) of the main text, its symplectic eigenvalues are equal

$$\nu_\pm = \sqrt{u\left(u - \frac{v^2}{w}\right)}, \tag{B4}$$

and the entropy of the entire system is

$$S(\hat{\rho}_{mm}) = 2g\left(\sqrt{u\left(u - \frac{v^2}{w}\right)}\right). \tag{B5}$$

The RCI is thus

$$I_R(\hat{\rho}_{mm}) = g\left(u - \frac{v^2}{2w}\right) - 2g\left(\sqrt{u\left(u - \frac{v^2}{w}\right)}\right). \tag{B6}$$

On the other hand, EOF quantifies the resources that are required to create a quantum state in terms of the number of Bell pairs (or "ebit"), which is formally defined as [24]

$$E_f(\hat{\rho}) = \min_i \sum_i p_i S(|\psi_i\rangle), \tag{B7}$$

where the minimum is taken over all possible pure state decomposition of state $\hat{\rho}$ as $\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. Specifically, for a pure state, both RCI and EOF reduce to entanglement entropy.

EOF is in general subadditive,

$$E_f\left(\hat{\rho}_1 \otimes \hat{\rho}_2\right) \leq E_f\left(\hat{\rho}_1\right) + E_f\left(\hat{\rho}_2\right). \tag{B8}$$

EOF is known to be analytically solvable for the symmetric two-mode Gaussian state $\hat{\rho}$ with the covariance matrix $V$ in Eq. (B2) when $a = b$ [58] as

$$E_f(\hat{\rho}) = \begin{cases} h(\sqrt{\nu_1 \nu_2}) & \text{if } \nu_1 \nu_2 < 1, \\ 0 & \text{otherwise,} \end{cases} \tag{B9}$$

where $\nu_1, \nu_2$ are the first two eigenvalues of $V$ in increasing order and $h(x)$ is defined as

$$h(x) \equiv \frac{(1+x)^2}{4x} \log_2\left(\frac{(1+x)^2}{4x}\right) - \frac{(1-x)^2}{4x} \log_2\left(\frac{(1-x)^2}{4x}\right). \tag{B10}$$

For the noisy entangled microwave modes [see Eq. (4) of the main text], we have $\nu_1 = \nu_2 = u - v^2/w$ and EOF is

$$E_f(\hat{\rho}_{m,m}) = h\left(u - \frac{v^2}{w}\right). \tag{B11}$$

EOF is analytically solvable for arbitrary two-qubit states [59] and we explain it in the following. We first introduce the spin-flipped state of $\hat{\rho}$ as

$$\hat{\tilde{\rho}} = (\hat{\sigma}_y \otimes \hat{\sigma}_y)\hat{\rho}^*(\hat{\sigma}_y \otimes \hat{\sigma}_y), \tag{B12}$$

where $\hat{\rho}^*$ is the complex conjugate of $\hat{\rho}$. Define the function $\mathcal{E}(x)$ as

$$\mathcal{E}(x) = H\left(\frac{1+\sqrt{1-x^2}}{2}\right), \tag{B13}$$

where $H(x) \equiv -x \log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy. Note that the $\mathcal{E}$ function monotonically increases with $x$ from 0 to 1. The EOF for an arbitrary two-qubit state $\hat{\rho}$ is thus

$$E_f(\hat{\rho}) = \mathcal{E}\left[\delta\left(\hat{\rho}\right)\right], \tag{B14}$$

where $\delta(\hat{\rho}) \equiv \max\{0, v_1 - v_2 - v_3 - v_4\}$ and $\{v_i\}_{i=1}^4$ are the square root of eigenvalues of $\hat{\rho}\hat{\tilde{\rho}}$ in decreasing order.

In the main text, we evaluate the entanglement measures for the noisy two-mode squeezed vacuum in Eq. (4) of the main text. When $\zeta_o = \zeta_m = 1$, $\hat{\rho}_{m,m}$ is in a pure TMSV state with mean photon number $N_{\text{Ideal}} = 16C^2/[(1-C)^2(1+6C+C^2)]$. In this pure state case, the EOF and RCI are both equal to $g\,(2N_{\text{Ideal}}+1)$.

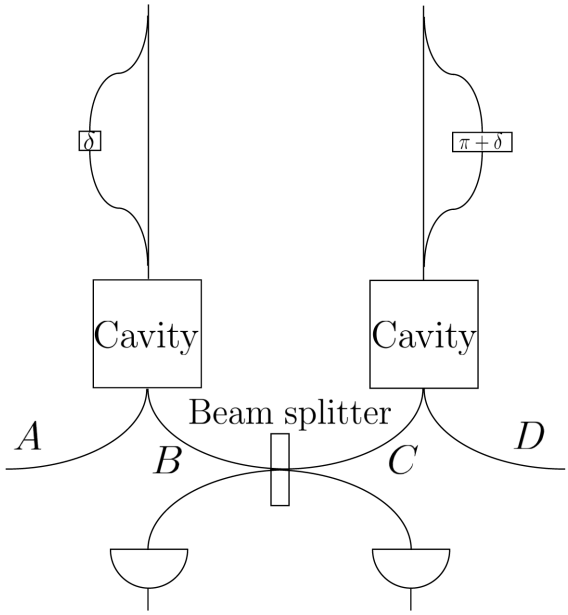


FIG. 8. Time-bin-encoding scheme. The pulses first go through preparation devices with the same path difference and different phase shifts. The time-bin-encoded quantum states are generated by the cavities. Then Alice and Bob send photons to the middle for entanglement swapping. Photon detection in the middle heralds the Bell state between Alice and Bob.

## APPENDIX C: TIME-BIN ENTANGLEMENT SWAP

We consider the single-photon-based protocol as shown in Fig. 8. Alice and Bob produce the pump pulses by preparation devices with the same path difference and phase shifts $\delta$ and $\delta + \pi$. Due to the path difference, pump pulse states are in superposition of early and late time bin. The pump pulse at Alice's side $|\psi\rangle_{P_1} = 1/\sqrt{2}(|e\rangle + e^{i\delta}|l\rangle)$ while at Bob's side $|\psi\rangle_{P_2} = 1/\sqrt{2}(|e\rangle - e^{i\delta}|l\rangle)$, where $|e\rangle$ and $|l\rangle$ are the "early" state and "late" state of a single photon, respectively. Then the input pump pulses are sent to the cavities to generate $M$-$O$ pairs. We assume a weak pump, then an output M-O pair can be generated from one cavity with probability $P_{11} = \text{Tr}(\hat{\rho}_{m,o}|11\rangle\langle 11|)$, where $\hat{\rho}_{m,o}$ is the Gaussian state with covariance matrix given by Eq. (2) of the main text.

We calculate the probability $P_{11}$ by integrating Wigner functions

$$\begin{aligned}&\text{Tr}(\hat{\rho}_{m,o}|11\rangle\langle 11|) \\ &\quad = (2\pi)^2 \int d^4x_1 x_2 p_1 p_2 W(x_1, p_1, x_2, p_2; \hat{\rho}_{m,o}) \\ &\qquad \times W(x_1, p_1, x_2, p_2; |11\rangle\langle 11|).\end{aligned} \tag{C1}$$

The Wigner function of Fock state $|n\rangle\langle n|$ is

$$W_n(x,p) = \frac{1}{\pi}(-1)^n \exp -(x^2+p^2)\, L_n\left(2(x^2+p^2)\right), \tag{C2}$$

where $L_n$ is Laguerre polynomial. Substituting this into the above equation we get

$$\begin{aligned} P_{11} &= \mathrm{Tr}(\hat{\rho}_{m,o}|11\rangle\langle 11|) \\ &= \frac{4(1+6v^2+v^4-2uwv^2-w^2-u^2+u^2w^2)}{(1+u-v^2+w+uw)^3}. \end{aligned} \tag{C3}$$

Similarly, we also obtain the probabilities of getting zero photon in microwave domain and one photon in optical domain

$$P_{01} = \mathrm{Tr}(\hat{\rho}_{m,o}|01\rangle\langle 01|) = \frac{4(-1-u-v^2+w+uw)}{(1+u-v^2+w+uw)^2}, \tag{C4}$$

one photon in microwave domain and zero photon in optical domain

$$P_{10} = \mathrm{Tr}(\hat{\rho}_{m,o}\,|10\rangle\,\langle 10|) = \frac{4(-1+u-v^2-w+uw)}{(1+u-v^2+w+uw)^2}, \tag{C5}$$

and zero photon in both

$$P_{00} = \mathrm{Tr}(\hat{\rho}_{m,o}\,|00\rangle\,\langle 00|) = \frac{4}{1+u-v^2+w+uw}. \tag{C6}$$

Note that the weak pump region is when the cooperativity $C \ll 1$. Figure 9 shows the probabilities versus the cooperativity. In general, one needs to consider the generation of two or more photons in a single temporal mode in the large cooperativity region; however, such events will be beyond the consideration of this paper and will in general make the performance worse, as those events lead to states out of the single-photon Hilbert space.

Below we analyze the case where single-photon pairs are generated successfully for both Alice and Bob. At Alice's and Bob's side the $M$-$O$ pairs are

$$|\Phi^+(\delta)\rangle_{AB} = \frac{1}{\sqrt{2}}(|e,e\rangle + e^{i\delta}\,|l,l\rangle), \tag{C7}$$

and

$$|\Phi^-(\delta)\rangle_{CD} = \frac{1}{\sqrt{2}}(|e,e\rangle - e^{i\delta}\,|l,l\rangle), \tag{C8}$$

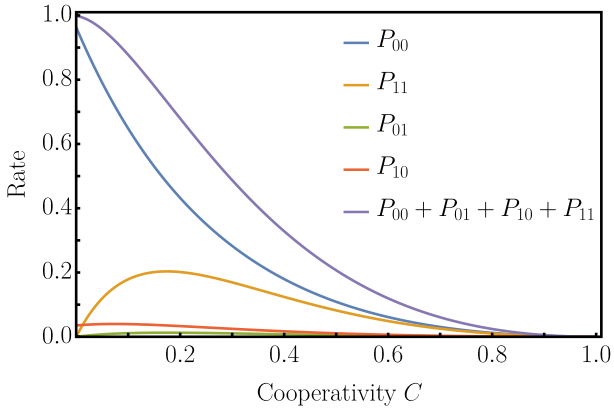


FIG. 9. The probability versus the cooperativity for $\zeta_o = 0.9$, $\zeta_m = 0.95$, and $n_{\mathrm{in}} = 0.2$.

respectively. The overall prepared state is in a product [60]

$$\begin{aligned} |\Phi^+(\delta)\rangle_{AB} \otimes |\Phi^-(\delta)\rangle_{CD} = \frac{1}{2}\big[&|\Phi^+\rangle_{BC} \otimes |\Phi^-(2\delta)\rangle_{AD} \\ &+ |\Phi^-\rangle_{BC} \otimes |\Phi^+(2\delta)\rangle_{AD} + |\Psi^+\rangle_{BC} \otimes e^{i\delta}\,|\Psi^-\rangle_{AD} \\ &+ |\Psi^-\rangle_{BC} \otimes e^{i\delta}\,|\Psi^+\rangle_{AD}\big], \end{aligned} \tag{C9}$$

where four Bell states are $|\Phi^\pm(\delta)\rangle = 1/\sqrt{2}(|e,e\rangle \pm e^{i\delta}\,|l,l\rangle)$ and $|\Psi^\pm(\delta)\rangle = 1/\sqrt{2}(|l,e\rangle \pm e^{i\delta}\,|e,l\rangle)$. Here $|\Phi^\pm\rangle$ and $|\Psi^\pm\rangle$ refer to the case of $\delta = 0$. Bell-state measurement in the middle consists of a balanced beam splitter and two photon detectors. The detection process is as follows. To detect

$$\begin{aligned} |\Psi^+\rangle_{BC} &= \frac{1}{\sqrt{2}}(|e,l\rangle + |l,e\rangle), \\ &= \frac{1}{\sqrt{2}}(\hat{a}_B^{(1)\dagger}\hat{a}_C^{(2)\dagger} + \hat{a}_B^{(2)\dagger}\hat{a}_C^{(1)\dagger})\,|0\rangle\,, \end{aligned} \tag{C10}$$

where $\hat{a}_{B/C}^{(i)\dagger}$ is the creation operator for port $B/C$, time bin $i$. The beam splitter transforms the bosonic modes by

$$\begin{pmatrix} \hat{a}_B \\ \hat{a}_C \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \hat{a}_{B'} \\ \hat{a}_{C'} \end{pmatrix}, \tag{C11}$$

where $B'/C'$ corresponds to the output.

Substituting Eq. (C11) into Eq. (C10) we get

$$|\Psi^+\rangle_{BC} \to \frac{1}{\sqrt{2}}(\hat{a}_{B'}^{(1)\dagger}\hat{a}_{B'}^{(2)\dagger} - \hat{a}_{C'}^{(1)\dagger}\hat{a}_{C'}^{(2)\dagger})\,|0\rangle\,. \tag{C12}$$

Similarly we have

$$|\Psi^-\rangle_{BC} \to \frac{1}{\sqrt{2}}(\hat{a}_{C'}^{(1)\dagger}\hat{a}_{B'}^{(2)\dagger} - \hat{a}_{B'}^{(1)\dagger}\hat{a}_{C'}^{(2)\dagger})\,|0\rangle\,. \tag{C13}$$

Detecting a single photon successively at one detector $B'$ and $C'$ measures $|\Psi^+\rangle_{BC}$ and heralds $|\Psi^-\rangle_{AD}$. Detecting

a single photon at both detectors but with a time delay measures $|\Psi^-\rangle_{BC}$ and heralds $|\Psi^+\rangle_{AD}$. The above two are success events in generating a Bell pair.

The other two Bell states transform as

$$|\Phi^\pm\rangle \to \frac{1}{2\sqrt{2}}\Big[(\hat{a}^{(1)\dagger}_{B'}\hat{a}^{(1)\dagger}_{B'} - \hat{a}^{(1)\dagger}_{C'}\hat{a}^{(1)\dagger}_{C'}) \pm(\hat{a}^{(2)\dagger}_{B'}\hat{a}^{(2)\dagger}_{B'} - \hat{a}^{(2)\dagger}_{C'}\hat{a}^{(2)\dagger}_{C'})\Big]|0\rangle\,. \tag{C14}$$

In these cases, we detect two photons at one of the detectors $B'$ and $C'$. Because there is no time difference between the detections, we cannot distinguish these two cases.

When $BC$ is projected to either $|\Psi^+\rangle_{BC}$ or $|\Psi^-\rangle_{BC}$, then $AD$ ends up with $e^{i\delta}|\Phi^-\rangle_{AD}$ or $e^{i\delta}|\Phi^+\rangle_{AD}$. The probability of success is limited to 50% with perfect detection.

Now we consider the other events that are considered as success, while entanglement is not generated. When both $B$ and $C$ have photons, it will trigger success events.

The first case is when $A$ gets zero photon, while $B$,$C$,$D$ all get one photon. This happens with probability $P_{01}P_{11}$. In this case, the states before swap are

$$|\Phi^+(\delta)\rangle_{AB} = \frac{1}{\sqrt{2}}(|0,e\rangle + e^{i\delta}|0,l\rangle), \tag{C15}$$

$$|\Phi^-(\delta)\rangle_{CD} = \frac{1}{\sqrt{2}}(|e,e\rangle - e^{i\delta}|l,l\rangle). \tag{C16}$$

Replacing $|0\rangle$ in the above analysis, after we have success detection, the state between $AD$ ends in

$$|\Psi^\pm\rangle_{AD} = \frac{1}{\sqrt{2}}(|0,e\rangle \pm |0,l\rangle), \tag{C17}$$

$$= |0\rangle_A \otimes \frac{1}{\sqrt{2}}(|e\rangle \pm |l\rangle)_D. \tag{C18}$$

Similarly, in the second case, when $D$ gets zero photon, and $A$,$B$,$C$ all have one photon, with probability also $P_{01}P_{11}$, we have the final state

$$|\Psi^\pm\rangle_{AD} = \frac{1}{\sqrt{2}}(|e\rangle \pm |l\rangle)_A \otimes |0\rangle_D\,. \tag{C19}$$

The final case is when $A$ and $D$ have zero photon, while $B$ and $C$ have one photon, which happens with probability $P_{01}^2$. In this case, we have

$$|\Psi^\pm\rangle_{AD} = |0,0\rangle\,, \tag{C20}$$

as a tensor product of vacuum.

So the "success" event happens with probability

$$P_{\text{success}} = (P_{11}^2 + 2P_{01}P_{11} + P_{01}^2)/2. \tag{C21}$$

The overall state conditioned on success is equivalent to the following state:

$$\hat{\rho}_{AD} = \frac{1}{P_{11}^2 + 2P_{01}P_{11} + P_{01}^2} \Big[P_{11}^2\,|\text{Bell}\rangle\langle\text{Bell}| + P_{01}P_{11}\,|0\rangle\langle 0|\otimes|+\rangle\langle+| + P_{01}P_{11}\,|+\rangle\langle+|\otimes|0\rangle\langle 0| + P_{01}^2\,|00\rangle\langle 00|\Big], \tag{C22}$$

$$\equiv \lambda_1\,|\text{Bell}\rangle\langle\text{Bell}| + \lambda_2\,|0\rangle\langle 0|\otimes|+\rangle\langle+| + \lambda_2\,|+\rangle\langle+|\otimes|0\rangle\langle 0| + \lambda_3\,|00\rangle\langle 00|\,, \tag{C23}$$

where we define $\lambda_1, \lambda_2, \lambda_3$ implicitly. The entanglement generation rate is therefore

$$R(\hat{\rho}_{AD}) = E(\hat{\rho}_{AD})P_{\text{success}}/2, \tag{C24}$$

where we divide by 2 since we use the channel twice for time-bin encoding and the $E()$ can be the RCI or EOF for entanglement measure.

The RCI of $\hat{\rho}_{AD}$ is evaluated as the following. First the entropy

$$S(\hat{\rho}_{AD}) = -\lambda_1\log_2(\lambda_1) - 2\lambda_2\log_2(\lambda_2) - \lambda_3\log_2(\lambda_3). \tag{C25}$$

The reduced state

$$\hat{\rho}_A = \text{Tr}_D(\hat{\rho}_{AD}), = \left(\frac{\lambda_1}{2} + \frac{\lambda_2}{2}\right)|l\rangle\langle l| + \left(\frac{\lambda_1}{2} + \frac{\lambda_2}{2}\right)|e\rangle\langle e| + \frac{1}{2}\lambda_2|e\rangle\langle l| + \frac{1}{2}\lambda_2|l\rangle\langle e| + (\lambda_2+\lambda_3)\,|0\rangle\langle 0|\,. \tag{C26}$$

The three eigenvalues of $\hat{\rho}_A$ are

$$\lambda_1' = \frac{1}{2}\lambda_1,$$
$$\lambda_2' = \frac{1}{2}(\lambda_1 + 2\lambda_2),$$
$$\lambda_3' = \lambda_2 + \lambda_3.$$

Therefore, the RCI is

$$S\left(\hat{\rho}_A\right) - S\left(\hat{\rho}_{AD}\right) = -\sum_{i=1}^{3}\lambda_i'\log_2(\lambda_i') - S\left(\hat{\rho}_{AD}\right). \tag{C27}$$

However, it is hard to evaluate EOF of $\hat{\rho}_{AD}$ and thus we consider its upper bound by the fidelity to Bell pair

$$F_{S\text{-}P} = \frac{P_{11}^2}{P_{11}^2 + 2P_{01}P_{11} + P_{01}^2}. \tag{C28}$$

Following Eq. (B7)

$$E_f(\hat{\rho}_{AD}) \leq \frac{P_{11}^2}{P_{11}^2 + 2P_{01}P_{11} + P_{01}^2} S(|\text{Bell}\rangle) = F_{S\text{-}P}, \tag{C29}$$

as the pure-state mixture in Eq. (C23) is orthorgonal.

## APPENDIX D: OPTICAL TRANSMISSION LOSS

We consider identical pure loss channels with transmissivity $\eta$ before the beam splitter. The channel is described by

$$\hat{a} \to \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{e}, \tag{D1}$$

where $\hat{a}$ and $\hat{e}$ represent the optical mode and the environmental vacuum separately. In the CV entanglement swap case, the parameters under the loss are transformed as

$$u \to u = 1 + \frac{8\zeta_m[C + n_{\text{in}}(1-\zeta_m)]}{(1-C)^2}, \tag{D2a}$$

$$v \to \sqrt{\eta}v = \frac{4\sqrt{\zeta_o\eta\zeta_m C}[1 + C + 2n_{\text{in}}(1-\zeta_m)]}{(1-C)^2}, \tag{D2b}$$

$$w \to (1-\eta) + \eta w = 1 + \frac{8C\zeta_o\eta\,[1 + n_{\text{in}}\,(1-\zeta_m)]}{(1-C)^2}. \tag{D2c}$$

One can easily check that, transmitting the noisy entangled CV modes through a pure loss channel of transmissivity $\eta$ is equivalent to the state with $\zeta_o \to \eta\zeta_o$, via Eq. (3a) in the main text. When $\eta \leq 0.5$ (equivalent to $\zeta_o \leq 0.5$), the CV entanglement swap enabled rate is zero, as the pure loss channel below half transmissivity has zero rate. In Fig. 10, we see that our approach allows an advantage when $\zeta_o \gtrsim 0.5$, which corresponds to 15 kilometers of state-of-the-art fiber link to the center swap node.

In the time-bin entanglement case, the analysis is similar, simply replacing $\zeta_o \to \eta\zeta_o$ in previous expressions of Sec. C. Note that one cannot simply multiply the success probability by $\eta^2$, as two or more photons can lead to a single photon after loss.

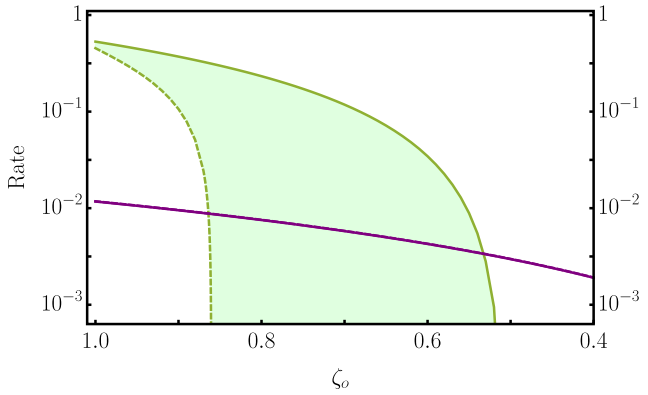


FIG. 10. Rate versus $\zeta_o$ when $\zeta_m = 0.992$, $n_{\text{in}} = 0.2$, and $C = 0.1$. Solid and dashed green curves represent lower and upper bounds for rate of CV entanglement swap. Purple line shows the rate of time-bin entanglement swap.

## APPENDIX E: SOLVING THE DIRECT SWAP

In this section, we provide more details on the direct swap approach, for entanglement distillation on the hybrid CV-DV platform. For a given pair of noisy entangled microwave modes in state $\hat{\rho}_{m,m}$, the composite system is prepared same as it is in the hybrid LVQC approach where two qubits are in state $|0\rangle_q$, shown in the top and bottom halves of Fig. 11(a). The CV-DV systems on both sides are evolved separately by the same unitary $\exp(-it\hat{H}_{\text{swap},\ell})$ where $\ell = 1, 2$ stands for the two sides. Through the evolution, we monitor the entanglement (i.e., EOF) between two qubits, as shown in Fig. 11(b). We stop the evolution and discard the modes when the two qubits are maximally entangled at the time $t = (2n+1)\pi/2, n \in \mathbb{N}$.

In the direct swap approach, we choose the final two-qubit state to maximize the entanglement, whose fidelity to $|\Psi^+\rangle$ is not guaranteed to be the maximal. To maximize the fidelity, we allow an arbitrary local unitary $\hat{U}$ on one of the qubits. Note that the Bell state $|\Psi^+\rangle$ is invariant under $\hat{U} \otimes \hat{U}^\star$, so all local unitary can be absorbed into a single qubit. We parameterize the single-qubit unitary by three angles and numerically maximize the fidelity to obtain the results.

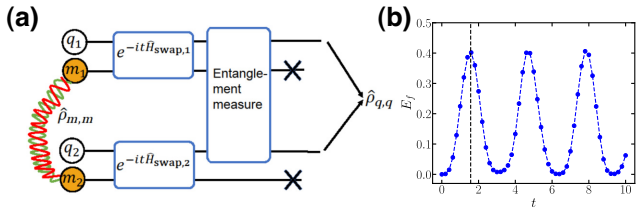


FIG. 11. (a) Schematic of the direct swap approach. (b) Entanglement of formation $E_f$ between qubits versus evolution time $t$. Vertical dashed line indicates the time with maximum entanglement at $t = \pi/2$.

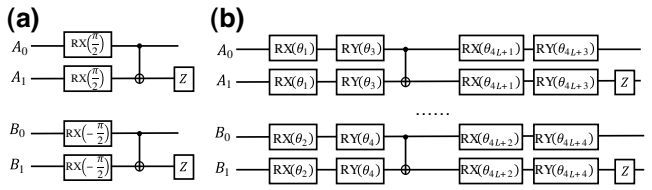


FIG. 12. Schematic of (a) DEJMPS protocol and (b) $L$-layer DV LVQC.

## APPENDIX F: QUBIT DISTILLATION PROTOCOLS

As explained in the main text, to obtain states with further lower infidelity, we consider distillation protocols on two copy of the output qubits produced from hybrid LVQCs or direct swap approach. Various entanglement distillation protocols on DV systems have been proposed, including BBPSSW [25], DEJMPS [26], and LOCCNet [28]. We mainly focus on the DEJMPS protocol, which is proved to be optimal for bell-diagonal states with rank up to three, and a DEJMPS-inspired DV LVQC, shown in Figs. 12(a) and 12(b) separately. For both protocols, the two-qubit system $A_0, B_0$ and $A_1, B_1$ are separately initialized with identical two-qubit mixed state produced from a hybrid distillation approach. Through the circuit, qubits $A_1, B_1$ are measured in the computational basis where only $|00\rangle$ or $|11\rangle$ are considered as success. The DV LVQC is trained in a similar way to the way we train the hybrid LVQC described in our main text and details can be found in Figs. 14(c) and 14(d).

## APPENDIX G: NUMERICAL DETAILS

In this section, we provide more details on the numerical simulation of the hybrid LVQC approach. For each choice of hyperparameters $\lambda, F_c$, we start with a batch of 500 instances with random initialization. The quantum circuits are implemented in PyTorch [61] and optimized by Adam [62] with the learning rate LR $= 0.001$ in 20 000 steps. Unless further specified, in this section we utilize hybrid LVQCs with $D = 10$ gate blocks and DV LVQCs with $L = 6$ layers. In the distillation, we choose postselection on photon number below $D/2$ as success.

### 1. Details of Fig. 3 of the main text

To begin with, we analyze the effect of circuit depth of the hybrid LVQC. We consider the same setting as Fig. 3(a) of the main paper, but choose different LVQC depth $D$ and the results are in Fig. 13. As we see, in the low infidelity region, the performance does improve as the depth increases. Indeed, a larger depth $D \geq 10$ can potentially further improve the performance, and further enlarge the advantage.

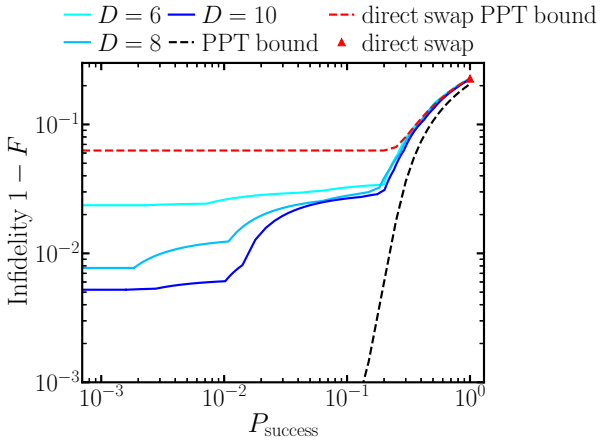


FIG. 13. Infidelity versus success probability of one-copy noisy entangled microwave modes with $C = 0.1, n_{\text{in}} = 0.2, \zeta_m = 0.992, \zeta_o = 0.99$ by utilizing hybrid LVQC with $D$ ECD blocks.

Now we present more details on the training results utilizing hybrid LVQCs. In Figs. 14(a) and 14(b), we show the results of the one-copy case, and the two-copy DEJMPS case. To fulfill the whole range of success probability $P_{\text{success}} \in [0, 1]$, we adopt different combinations of hyperparameters $\lambda, F_c$. To obtain the best performance from hybrid LVQCs in the two-copy scenario, we extract the instances with highest fidelity within each small bin of success probability and apply interpolation (black solid curves) between them. The details of the interpolation are introduced in the next paragraph. In Fig. 14(c), we show training results with DV LVQC [see Fig. 12(b) for the circuit design] on those best output qubit pairs in (a) with various combinations of $\lambda, F_c$. Note that there is advantages from DV LVQCs in terms of infidelity compared to DEJMPS, and thus the combination of the DEJMPS and DV LVQCs provides the best DV approach results for the two-copy case, shown in Fig. 3(b) in the main text. In Fig. 14(d), we apply the same DV LVQC approach on the two-qubit state generated by direct swap. To summarize, compared to the DEJMPS protocol, the DV LVQC approach can provide better performance with smaller infidelity at the expense of more quantum gates.

To obtain the best performance of variational circuits for each given success probability, we utilize interpolation and extrapolation [27]. For any two protocols with success probability $p_1, p_2$ and fidelity $F_1, F_2$, we consider the probabilistic mixing of the protocols with probability $r, 1 - r \in [0, 1]$, then the success probability and fidelity for the protocol mixture are

$$P_{\text{success,int}} = rp_1 + (1 - r)p_2, \tag{G1a}$$

$$F_{\text{int}} = \frac{rp_1F_1 + (1 - r)p_2F_2}{P_{\text{success,int}}}. \tag{G1b}$$

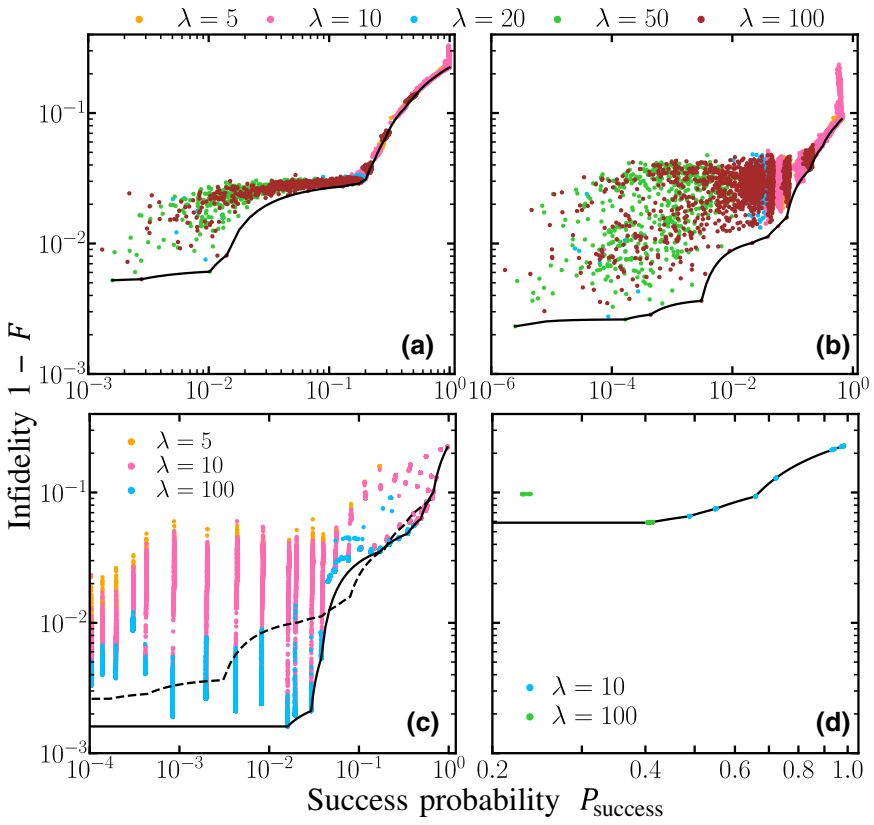


FIG. 14. Numerical simulation details of infidelity versus success probability with (a) hybrid LVQC; (b) hybrid LVQC+DEJMPS; (c) hybrid LVQC+DV LVQC; (d) direct swap+DV LVQC. Scatter dots with different color represent training results with different choice of hyperparameters $\lambda, F_c$ and random initializations. Black solid curves show the smallest infidelity for given success probability for each approach with interpolation. Black dashed curve in (c) represents the smallest infidelity for hybrid LVQC+DEJMPS [same as the black solid curve in (b)].

Similarly, one can always extrapolate a protocol to a trivial strategy that always output a product of $|0\rangle_q$ state, with an identity success probability but $1/2$ fidelity. The protocol with extrapolation has success probability and fidelity as

$$P_{\text{success,ext}} = rp_1 + (1-r), \tag{G2a}$$

$$F_{\text{ext}} = \frac{rp_1F_1 + (1-r)/2}{P_{\text{success,ext}}}. \tag{G2b}$$

### 2. Details of Fig. 4 of the main text

In Fig. 15, we provide the details of the rate versus infidelity plot of Fig. 4 in the main text. We present the data points from different training, under various hyperparameters indicated by the different coloring. Subplot (a) presents the single-copy case, while (b) and (c) present the two-copy case, with the DEJMPS and DV LVQC separately. The black solid lines in (b) and (c) are the best performance of the results in both (b) and (c), which are eventually presented in Fig. 4(b) of the main text.

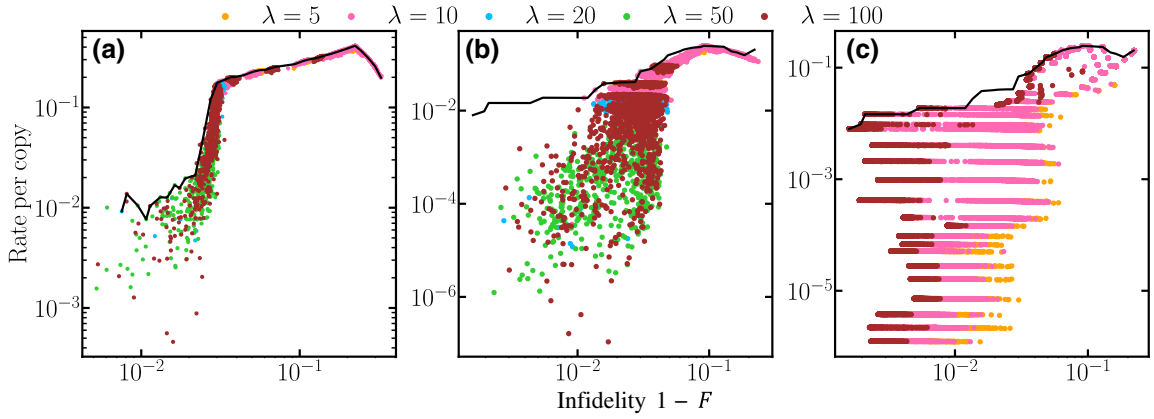


FIG. 15. Numerical simulation details of distillation rate $R_{E_f}$ versus infidelity for (a) one-copy hybrid LVQC, (b),(c) two-copy hybrid LVQC+DEJMPS and hybrid LVQC+DV LVQC. Scatter dots with different color represent training results with different choice of hyperparameters $\lambda, F_c$ and random initializations. Black solid curves show the highest rate given infidelity for (a) one copy and (b),(c) two copy.

[1] P. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[2] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[3] L. K. Grover, in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (1996), p. 212.

[4] I. Bloch, J. Dalibard, and S. Nascimbene, Quantum simulations with ultracold quantum gases, Nat. Phys. **8**, 267 (2012).

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, and D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[6] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, and D. Fan, *et al.*, Strong Quantum Computational Advantage using a Superconducting Quantum Processor, Phys. Rev. Lett. **127**, 180501 (2021).

[7] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, and R. J. Schoelkopf, *et al.*, Quantum error correction of a qubit encoded in grid states of an oscillator, Nature **584**, 368 (2020).

[8] A. Eickbusch, V. Sivak, A. Z. Ding, S. S. Elder, S. R. Jha, J. Venkatraman, B. Royer, S. Girvin, R. J. Schoelkopf, and M. H. Devoret, Fast universal control of an oscillator with weak dispersive coupling to a qubit, Nat. Phys. (2022).

[9] H. J. Kimble, The quantum internet, Nature **453**, 1023 (2008).

[10] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[11] W. Kozlowski and S. Wehner, in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication* (2019), p. 1.

[12] P. Magnard, S. Storz, P. Kurpiers, J. Schär, F. Marxer, J. Lütolf, T. Walter, J.-C. Besse, M. Gabureac, K. Reuer, A. Akin, B. Royer, A. Blais, and A. Wallraff, Microwave Quantum Link between Superconducting Circuits Housed in Spatially Separated Cryogenic Systems, Phys. Rev. Lett. **125**, 260502 (2020).

[13] Y. Zhong, H.-S. Chang, A. Bienfait, É. Dumur, M.-H. Chou, C. R. Conner, J. Grebel, R. G. Povey, H. Yan, and D. I. Schuster, *et al.*, Deterministic multi-qubit entanglement in a quantum network, Nature **590**, 571 (2021).

[14] X. Han, W. Fu, C.-L. Zou, L. Jiang, and H. X. Tang, Microwave-optical quantum frequency conversion, Optica **8**, 1050 (2021).

[15] R. W. Andrews, R. W. Peterson, T. P. Purdy, K. Cicak, R. W. Simmonds, C. A. Regal, and K. W. Lehnert, Bidirectional and efficient conversion between microwave and optical light, Nat. Phys. **10**, 321 (2014).

[16] J. Bochmann, A. Vainsencher, D. D. Awschalom, and A. N. Cleland, Nanomechanical coupling between microwave and optical photons, Nat. Phys. **9**, 712 (2013).

[17] L. Fan, C.-L. Zou, R. Cheng, X. Guo, X. Han, Z. Gong, S. Wang, and H. X. Tang, Superconducting cavity electro-optics: A platform for coherent photon conversion between superconducting and photonic circuits, Sci. Adv. **4**, eaar4994 (2018).

[18] Y. Xu, A. A. Sayem, L. Fan, S. Wang, R. Cheng, C.-L. Zou, W. Fu, L. Yang, M. Xu, and H. X. Tang, Bidirectional electro-optic conversion reaching 1% efficiency with thin-film lithium niobate, (2020) ArXiv:2012.14909.

[19] R. Hisatomi, A. Osada, Y. Tabuchi, T. Ishikawa, A. Noguchi, R. Yamazaki, K. Usami, and Y. Nakamura, Bidirectional conversion between microwave and light via ferromagnetic magnons, Phys. Rev. B **93**, 174427 (2016).

[20] L. A. Williamson, Y.-H. Chen, and J. J. Longdell, Magneto-optic Modulator with Unit Quantum Efficiency, Phys. Rev. Lett. **113**, 203601 (2014).

[21] J. G. Bartholomew, J. Rochman, T. Xie, J. M. Kindem, A. Ruskuc, I. Craiciu, M. Lei, and A. Faraon, On-chip coherent microwave-to-optical transduction mediated by ytterbium in $YVO_4$, Nat. Commun. **11**, 3266 (2020).

[22] S. Krastanov, H. Raniwala, J. Holzgrafe, K. Jacobs, M. Lončar, M. J. Reagor, and D. R. Englund, Optically Heralded Entanglement of Superconducting Systems in Quantum Networks, Phys. Rev. Lett. **127**, 040503 (2021).

[23] J. Wu, C. Cui, L. Fan, and Q. Zhuang, Deterministic Microwave-Optical Transduction based on Quantum Teleportation, Phys. Rev. Appl. **16**, 064044 (2021).

[24] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[25] 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).

[26] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Quantum Privacy Amplification and the Security of Quantum Cryptography over Noisy Channels, Phys. Rev. Lett. **77**, 2818 (1996).

[27] F. Rozpędek, T. Schiet, D. Elkouss, A. C. Doherty, and S. Wehner, *et al.*, Optimizing practical entanglement distillation, Phys. Rev. A **97**, 062333 (2018).

[28] X. Zhao, B. Zhao, Z. Wang, Z. Song, and X. Wang, Practical distributed quantum information processing with LOCCNet, npj Quantum Inf. **7**, 159 (2021).

[29] J.-W. Pan, C. Simon, Č. Brukner, and A. Zeilinger, Entanglement purification for quantum communication, Nature **410**, 1067 (2001).

[30] P. G. Kwiat, S. Barraza-Lopez, A. Stefanov, and N. Gisin, Experimental entanglement distillation and 'hidden' nonlocality, Nature **409**, 1014 (2001).

[31] T. Yamamoto, M. Koashi, Ş. K. Özdemir, and N. Imoto, Experimental extraction of an entangled photon pair from two identically decohered pairs, Nature **421**, 343 (2003).

[32] R. Reichle, D. Leibfried, E. Knill, J. Britton, R. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Experimental purification of two-atom entanglement, Nature **443**, 838 (2006).

[33] T. C. Ralph and A. Lund, in *AIP Conference Proceedings*, Vol. 1110 (American Institute of Physics, 2009), p. 155.

[34] D. T. Pegg, L. S. Phillips, and S. M. Barnett, Optical State Truncation by Projection Synthesis, Phys. Rev. Lett. **81**, 1604 (1998).

[35] H. Takahashi, J. S. Neergaard-Nielsen, M. Takeuchi, M. Takeoka, K. Hayasaka, A. Furusawa, and M. Sasaki, Entanglement distillation from Gaussian input states, Nat. Photonics **4**, 178 (2010).

[36] S. Zhang and P. van Loock, Local Gussian operations can enhance continuous-variable entanglement distillation, Phys. Rev. A **84**, 062309 (2011).

[37] A. Datta, L. Zhang, J. Nunn, N. K. Langford, A. Feito, M. B. Plenio, and I. A. Walmsley, Compact Continuous-Variable Entanglement Distillation, Phys. Rev. Lett. **108**, 060502 (2012).

[38] O. Černotík and J. Fiurášek, Displacement-enhanced continuous-variable entanglement concentration, Phys. Rev. A **86**, 052339 (2012).

[39] M. He, R. Malaney, and B. A. Burnett, Noiseless linear amplifiers for multimode states, Phys. Rev. A **103**, 012414 (2021).

[40] L. Hu, Z. Liao, and M. S. Zubairy, Continuous-variable entanglement via multiphoton catalysis, Phys. Rev. A **95**, 012310 (2017).

[41] Y. Mardani, A. Shafiei, M. Ghadimi, and M. Abdi, Continuous-variable entanglement distillation by cascaded photon replacement, Phys. Rev. A **102**, 012407 (2020).

[42] E. T. Campbell, M. G. Genoni, and J. Eisert, Continuous-variable entanglement distillation and noncommutative central limit theorems, Phys. Rev. A **87**, 042330 (2013).

[43] A. E. Ulanov, I. A. Fedorov, A. A. Pushkina, Y. V. Kurochkin, T. C. Ralph, and A. Lvovsky, Undoing the effect of loss on quantum entanglement, Nat. Photonics **9**, 764 (2015).

[44] M. Tsang, Cavity quantum electro-optics, Phys. Rev. A **81**, 063837 (2010).

[45] M. Tsang, Cavity quantum electro-optics. II. Input-output relations between traveling optical and microwave fields, Phys. Rev. A **84**, 043845 (2011).

[46] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[47] R. García-Patrón, S. Pirandola, S. Lloyd, and J. H. Shapiro, Reverse Coherent Information, Phys. Rev. Lett. **102**, 210501 (2009).

[48] C. Zhong, Z. Wang, C. Zou, M. Zhang, X. Han, W. Fu, M. Xu, S. Shankar, M. H. Devoret, and H. X. Tang, *et al.*, Proposal for Heralded Generation and Detection of Entangled Microwave–Optical-Photon Pairs, Phys. Rev. Lett. **124**, 010511 (2020).

[49] C. Zhong, X. Han, and L. Jiang, Quantum transduction with microwave and optical entanglement, (2022) ArXiv:2202.04601.

[50] J. Agustí, Y. Minoguchi, J. M. Fink, and P. Rabl, Long-distance distribution of qubit-qubit entanglement using Gaussian-correlated photonic beams, Phys. Rev. A **105**, 062454 (2022).

[51] J. Hastrup, K. Park, J. B. Brask, R. Filip, and U. L. Andersen, Universal Unitary Transfer of Continuous-Variable Quantum States into a Few Qubits, Phys. Rev. Lett. **128**, 110503 (2022).

[52] R. W. Heeres, B. Vlastakis, E. Holland, S. Krastanov, V. V. Albert, L. Frunzio, L. Jiang, and R. J. Schoelkopf, Cavity State Manipulation using Photon-Number Selective Phase Gates, Phys. Rev. Lett. **115**, 137002 (2015).

[53] S. Krastanov, V. V. Albert, C. Shen, C.-L. Zou, R. W. Heeres, B. Vlastakis, R. J. Schoelkopf, and L. Jiang, Universal control of an oscillator with dispersive coupling to a qubit, Phys. Rev. A **92**, 040303 (2015).

[54] C. Cui, C. N. Gagatsos, S. Guha, and L. Fan, High-purity pulsed squeezing generation with integrated photonics, Phys. Rev. Res. **3**, 013199 (2021).

[55] V. Vedral and M. B. Plenio, Entanglement measures and purification procedures, Phys. Rev. A **57**, 1619 (1998).

[56] F. G. Brandao and M. B. Plenio, Entanglement theory and the second law of thermodynamics, Nat. Phys. **4**, 873 (2008).

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

[58] S. Tserkis, J. Thompson, A. P. Lund, T. C. Ralph, P. K. Lam, M. Gu, and S. M. Assad, Maximum entanglement of formation for a two-mode Gaussian state over passive operations, Phys. Rev. A **102**, 052418 (2020).

[59] W. K. Wootters, Entanglement of Formation of an Arbitrary State of Two Qubits, Phys. Rev. Lett. **80**, 2245 (1998).

[60] H. de Riedmatten, I. Marcikic, J. A. W. van Houwelingen, W. Tittel, H. Zbinden, and N. Gisin, Long-distance entanglement swapping with photons from separated sources, Phys. Rev. A **71**, 050302 (2005).

[61] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, *et al.*, Pytorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. **32**, (2019), https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[62] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, ICLR (2015).