Editors' Suggestion

Number-Resolved Photocounter for Propagating Microwave Mode

R. Dassonneville,[1] R. Assouly,[1] T. Peronnin,[1] P. Rouchon,[2,3] and B. Huard[1,*]

[1] *Laboratoire de Physique, Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS, Lyon F-69342, France*
[2] *Centre Automatique et Systèmes, Mines-ParisTech, PSL Research University, 60 bd Saint-Michel, Paris 75006, France*
[3] *QUANTIC team, INRIA de Paris, 2 rue Simone Iff, Paris 75012, France*

(Received 10 April 2020; revised 4 September 2020; accepted 15 September 2020; published 14 October 2020)

Detectors of propagating microwave photons have recently been realized using superconducting circuits. However, a number-resolved photocounter is still missing. In this article, we demonstrate a single-shot counter for propagating microwave photons that can resolve up to three photons. It is based on a pumped Josephson ring modulator that can catch an arbitrary propagating mode by frequency conversion and store its quantum state in a stationary memory mode. A transmon qubit then counts the number of photons in the memory mode using a series of binary questions. Using measurement-based feedback, the number of questions is minimal and scales logarithmically with the maximal number of photons. The detector features a detection efficiency of 0.96 ± 0.04, and a dark-count probability of 0.030 ± 0.002 for an average dead time of 4.5 μs. To maximize its performance, the device is first used as an *in situ* waveform detector from which an optimal pump is computed and applied. Depending on the number of incoming photons, the detector succeeds with a probability that ranges from 54 ± 2 to 99%.

DOI: 10.1103/PhysRevApplied.14.044022

I. INTRODUCTION

Photon detectors are an important element in the quantum optics toolbox. At optical frequencies, detectors such as single-photon avalanche photodiodes or superconducting nanowire single-photon detectors are readily available [1]. In contrast, at GHz frequencies, these kinds of absorptive detectors are harder to realize due to the low energy of the microwave photons, roughly 5 orders of magnitude lower compared to their optical counterparts. Detecting and counting the microwave photons of a stationary mode is nowadays routinely performed using the dispersive interaction with a qubit [2–6]. These operations remain challenging for propagating photons because the light-matter interaction time is smaller. Yet some photon detectors for propagating modes have been proposed [7–17] and developed based on various approaches: direct absorption [18,19], encoding parity in the phase of a qubit [20,21], encoding the probability to have a single photon in a qubit excitation [22], or reservoir engineering [23]. Several implementations of a photocounter—a microwave photodetector able to resolve the photon number—for a propagating mode have been proposed [7,17,21,24,25]. However, such a device has yet to be demonstrated. Indeed, Refs. [20,21] only distinguish the parity of the photon number. References [22,23] only distinguish Fock state $|1\rangle$ from the rest while Refs. [18,19] distinguish 0 photon from at least 1.

Here, we demonstrate a photocounter that resolves the number of photons in a given propagating mode. To optimize the efficiency of our counter, we devise a way to calibrate *in situ* the arrival time and envelope of the propagating mode. The device can distinguish between 0, 1, 2, and 3 photons in a 20-MHz band around 10.220 GHz using measurement-based feedback. Finally, we propose a parameter-free model that accurately predicts the behavior of the counter, as demonstrated by coherent-state photocounting and Wigner tomography.

II. DEVICE AND OPERATION

The purpose of a photocounter is to count the photon number in a propagating mode with state $|\psi\rangle$ by providing an integer outcome n with probability $|\langle\psi|n\rangle|^2$. Our photocounter proceeds in three steps [Fig. 1(a)]. In step ①, it catches the incoming wavepacket and converts it into a high-Q stationary mode (memory). Then, in step ②, it counts the number of photons in the memory using an ancillary qubit. Finally (step ③), it resets the memory and qubit in their ground state. The catch and memory-reset operations (①, ③) are performed by frequency conversion using a Josephson ring modulator (JRM) [26,27]. The input transmission line is coupled to a buffer mode at frequency $\omega_b/2\pi = 10.220$ GHz, which sets the operating bandwidth of the counter to

[*] benjamin.huard@ens-lyon.fr

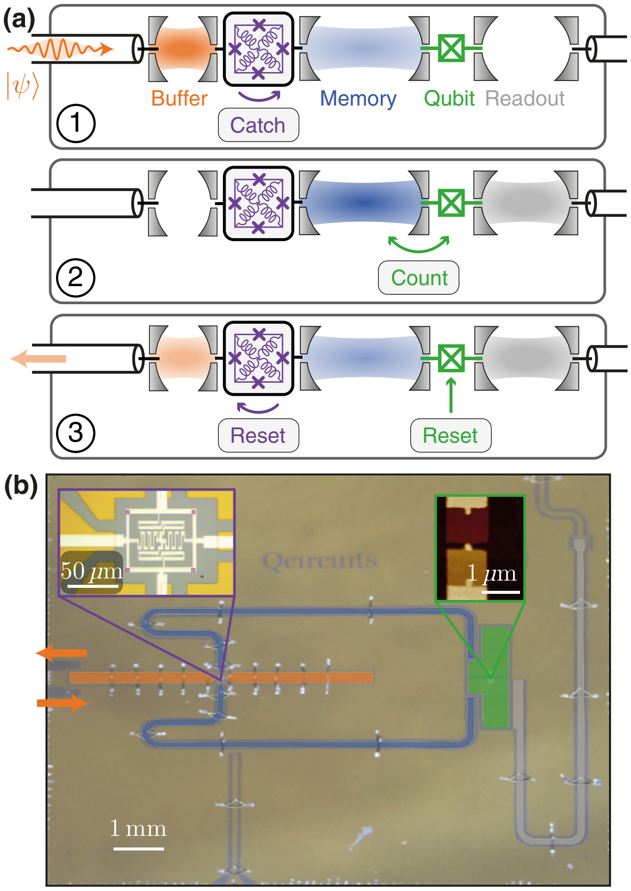

FIG. 1. Principle of operation and device. (a) A propagating microwave mode in state $|\psi\rangle$ is sent to the device via a buffer resonator. It is caught ① into the memory by pumping the Josephson ring modulator (JRM). The qubit then counts ② the photon number in the memory. The device is finally reset ③. Pumping the JRM empties the memory by releasing its photons into an arbitrary outgoing mode. The qubit is put into its ground state by measurement-based feedback. (b) False color image of the device where a JRM (left inset) is located at the crossing between buffer and memory $\lambda/2$ resonators. A transmon qubit (right inset) is coupled both to the memory and readout resonators.

$\kappa_b = 2\pi \times 20$ MHz $= (8.0 \text{ ns})^{-1}$. When pumped by a coherent tone of amplitude $p(t)$ at $\omega_b - \omega_m$, the JRM introduces a frequency-conversion term $\hat{H}_{\text{JRM}} = g_3 p(t)\hat{b}\hat{m}^\dagger +$ h.c. between the buffer $\hat{b}$ and the memory $\hat{m}$. The memory resonates at $\omega_m/2\pi = 3.74527$ GHz with a relaxation time $T_{1,m} = 4$ μs. When the memory is initially empty, this term enables us to catch the incoming wavepacket onto the buffer by storing its quantum state in the memory. Conversely, when the counting operation is over, we use it to release the photons from the memory into an arbitrary outgoing wavepacket.

From the point of view of the memory, the pumped JRM induces a tunable coupling to a transmission line [28]. It is thus possible to catch or release an arbitrary wavepacket into and from the memory [29–35]. Besides, the parasitic nonlinearities induced by the Josephson junctions of the JRM can be canceled by setting the flux through the JRM optimally, which we did (Appendix B). Using input-output formalism in the rotating frame, and neglecting the relaxation of the memory, the dynamics is captured by

$$\begin{aligned} \frac{d\hat{b}}{dt} &= -\frac{\kappa_b}{2}\hat{b} - g_3 p^*(t)\hat{m} + \sqrt{\kappa_b}\hat{b}_{\text{in}}(t), \\ \frac{d\hat{m}}{dt} &= g_3^* p(t)\hat{b}. \end{aligned} \tag{1}$$

For any given envelope $\langle b_{\text{in}}(t)\rangle$ of the incoming wavepacket that fits inside the buffer bandwidth κ_b, there exists an optimal pump $p_{\text{opt}}(t)$ for which the incoming quantum state is perfectly swapped into the memory [36]. For instance, if the incoming wavepacket is $\langle b_{\text{in}}(t)\rangle \propto 1/\cosh(\sqrt{\pi/2}t/\sigma)$ [(Fig. 5(a)], the optimal catching pump is given by $p_{\text{opt}}(t) \propto [1 + \lambda/2 \tanh(\lambda\kappa_b t/4)]\,(e^{\lambda\kappa_b t/2} + 1 - \lambda/2)^{-1/2}$. where $\lambda = \sqrt{8\pi}/\kappa_b\sigma$. Note that even at nonoptimal flux through the JRM or with finite relaxation time of the memory, an optimal pump can be found to catch the entire wavepacket (Appendix D).

III. BUILT-IN SAMPLE-AND-HOLD POWER METER

In order to generate the optimal pump $p_{\text{opt}}(t)$ for an arbitrary incoming wavepacket at ω_b, one needs to determine the envelope $\langle b_{\text{in}}(t)\rangle$. Interestingly, the envelope of any incoming waveform, can be determined *in situ*. The photocounter can indeed operate as a sample-and-hold power meter. Turning on the pump for a short sampling time of 20 ns after a variable delay t_d and counting the mean number of photons in the memory, using the coupled transmon qubit (Appendix E), enables us to directly probe $\langle b_{\text{in}}^\dagger b_{\text{in}}\rangle$ up to a global prefactor [Fig. 2(a)]. We demonstrate this functionality on a variety of generated waveforms displayed in Fig. 2(b) (left panel). The distortion of the waveforms introduced by the finite bandwidth κ_b of the counter and the nonzero sampling time can be seen in the measured mean photon number $\langle n\rangle$ as a function of t_d (right panel). The simple model Eq. (1) accurately reproduces the measured envelopes, where the only free parameter is the 15-ns difference in propagation time between buffer and pump lines.

IV. CATCH EFFICIENCY

In order to measure the catch efficiency η, we follow a catch-wait-release protocol as in Refs. [31,37]. We send an input signal $\langle b_{\text{in}}(t)\rangle \propto 1/\cosh(\sqrt{\pi/2}t/\sigma)$ of width $\sigma = 52$ ns and use the corresponding optimal pump shape, computed using Eq. (1). The calibration consists in measuring the outgoing amplitude b_{out} in various configurations. First, we measure the directly reflected amplitude

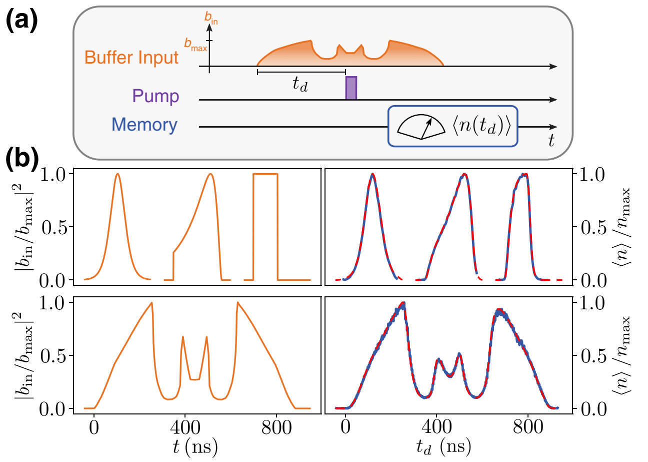

FIG. 2. *In situ* calibration of the incoming wavepacket envelope. (a) Amplitude of an arbitrary incoming wavepacket sent onto the buffer and of the sampling pump pulse. A following measurement of the mean photon number $\langle n(t_d)\rangle$ in the memory is performed using the qubit. (b) Left panels: various incoming waveforms. Right panels: solid blue (dashed red) lines show the measured [predicted using Eq. (1)] mean photon number $\langle n\rangle$ normalized by its maximum $n_{\max}$.

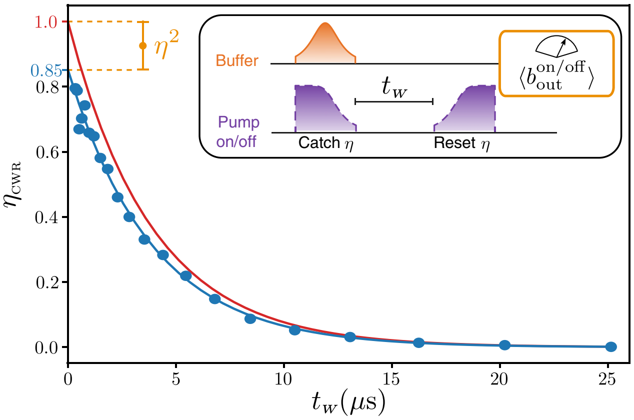

FIG. 3. Red solid line: round trip efficiency $\eta_{\rm CWR}$ as a function of the waiting time t_w assuming that the only imperfection comes from the memory decay with a characteristic time $T_{1,m}$. Blue dots: lower bound on the measured round trip efficiency $\eta_{\rm CWR}$. Blue solid line: exponential decay with characteristic time $T_{1,m}$. Orange error bar: range of possible values for η^2 that leads to a catch efficiency $\eta = 0.96 \pm 0.04$. Inset: pulse sequence of the catch-wait-release protocol. We measure the average outgoing amplitude $\langle b_{\rm out}^{\rm on/off}\rangle$ when the pump is on or off, from which we compute the round-trip efficiency $\eta_{\rm CWR} = \langle b_{\rm out}^{\rm on}\rangle^2/\langle b_{\rm out}^{\rm off}\rangle^2$.

$b_{\rm out}^{\rm off}$ without pumping, which provides a reference. Then, we measure the re-emitted amplitude $b_{\rm out}^{\rm on}$ after optimally catching, waiting a time t_w, and releasing. The round-trip efficiency is then given by $\eta_{\rm CWR} = \langle b_{\rm out}^{\rm on}\rangle^2/\langle b_{\rm out}^{\rm off}\rangle^2$. Besides, assuming that the catch and release operations have the same efficiency η, we get $\eta_{\rm CWR} = \eta^2 e^{-t_w/T_{1,m}}$, and thus an estimation of η.

In practice, due to the finite directivity of the directional coupler used to drive the buffer (Appendix A), there are interferences between the signal parasitically bypassing the coupler towards the output line and the desired signal coming from the buffer. This problem exclusively affects the denominator of the measured energy ratio since the parasitic signal does not spatially overlap with the signal that is released after t_w. In our case, the interferences are destructive, which leads to an underestimation of the denominator. As a consequence, we obtain apparent energy ratios in excess of 100%.

It is, however, possible to get a lower bound on the actual efficiency $\eta_{\rm CWR}(t_w)$ by measuring the coupler directivity. Right after the run, we measure a 16-dB directivity at room temperature using a calibrated vector network analyzer. In Fig. 3, the lowest possible values of $\eta_{\rm CWR}(t_w)$ (dots) are shown assuming fully destructive interferences in the denominator (correction by a factor 0.746 on the apparent energy ratio). Fitting these lower values by an exponential decreasing function at rate $1/T_{1,m}$, we get a lower bound on the catch efficiency $\eta = \sqrt{\eta_{\rm CWR}(t_w = 0)} \geq 0.92$.

V. BINARY DECOMPOSITION OF THE PHOTON NUMBER

Once the incoming wavepacket is characterized and efficiently caught, step ② consists in measuring the photon number present in the memory in a single-shot manner. To do so, we use a transmon qubit at frequency $\omega_q/2\pi = 4.32731$ GHz dispersively coupled to the memory such that $\hat{\mathcal{H}}_{qm} = -\chi\hat{m}^\dagger\hat{m}|e\rangle\langle e|$. Owing to a dispersive shift $\chi/2\pi = 3.28$ MHz much larger than the qubit decoherence rate $\Gamma_2 = (13.6\ \mu s)^{-1}$, the device operates in the photon-number-resolved regime [38]. It is thus possible to access information about the photon number by entangling the memory mode with the qubit and reading out its state. It is made possible by another resonator (readout), with frequency $\omega_r/2\pi = 6.293$ GHz, dispersively coupled to the qubit. We optimize the readout fidelity up to 97% in 252 ns, using a CLEAR-like sequence [39], mostly limited by the finite qubit relaxation time $T_1 = 7.1\ \mu s$ (Appendix C). The actual counting uses a scheme that measures the photon number bit by bit [40,41]. We denote u_k the kth least significant bit of $n = [u_N u_{N-1}\cdots u_1]_2$. Starting from u_1, each value of u_k is encoded into the qubit state and then read out. The main difficulty in implementing this scheme comes from the need to know the value of $n_{k-1} = [u_{k-1}\cdots u_1]_2$ in order to extract u_k. Each step Q_k [Fig. 4(a)] of the recursive determination of the u_k's is based on the relation

$$2^k u_k = n - n_{k-1} \bmod 2^k. \tag{2}$$

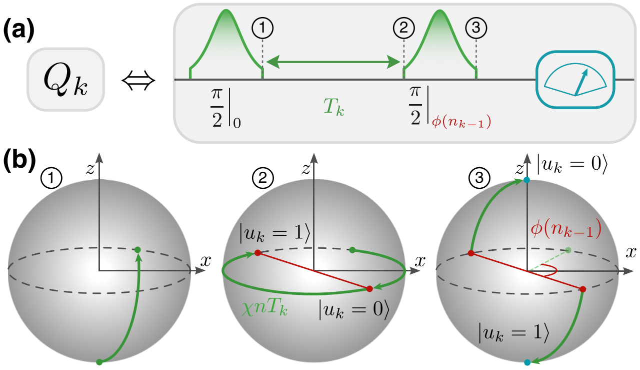

FIG. 4. Binary decomposition. (a) Pulse sequence used to extract u_k experimentally. Green corresponds to taking the remainder modulo 2^k of the photon number, red to the subtraction of the previously found digits $n_{k-1} = [u_{k-1} \cdots u_1]_2$ and blue to the extraction of the result via a measurement of the qubit. The $\pi/2$ pulses consist of sech waveforms with $\sigma = 4$ ns truncated at 4σ further optimized to mitigate the effect of the transmon qubit's low anharmonicity of -98 MHz [42]. (b) Trajectory of the qubit on the Bloch sphere when the cavity is in a Fock state $|n\rangle$ with yet unknown bit u_k. ①: the qubit is prepared in $(|g\rangle + i\,|e\rangle)/\sqrt{2}$ with an unconditional $\pi/2$ pulse around x. ②: trajectory of the qubit states $|u_k = 0\rangle$ and $|u_k = 1\rangle$ corresponding to the two possible values of the kth bit of the photon number during the waiting time T_k. Right ③: the last $\pi/2$ pulse around an axis shifted by an angle $\phi(n_{k-1})$ from the x axis maps u_k onto ground or excited states.

The qubit is prepared in $(|g\rangle + i\,|e\rangle)/\sqrt{2}$ with a $\pi/2$ pulse [Fig. 4(b) ①]. Then, the memory and qubit interact dispersively for a time $T_k = 2\pi/(\chi 2^k)$. T_k is chosen such that the qubit ends up in one of two orthogonal states $|u_k = 0\rangle$ and $|u_k = 1\rangle$ that only depend on the value u_k [Fig. 4(b) ②]. Precisely, the phase of the qubit states picks up an offset $\phi(n_{k-1}) = -n_{k-1} 2\pi/2^k$ for $u_k = 0$. Finally, using the knowledge of n_{k-1}, it is possible to map $|u_k = 0\rangle$ and $|u_k = 1\rangle$ onto $|e\rangle$ and $|g\rangle$ using a second $\pi/2$ pulse around the right axis [Fig. 4(b) ③]. Reading out the qubit state thus provides u_k directly. This scheme minimizes the number of qubit readouts required as each binary question Q_k is able to extract one bit of new information about the photon number. The number of binary questions required to determine a photon number n is $N = \lceil \log_2 n \rceil$, with the caveat that the necessary precision in the waiting time increases exponentially with N. Note that it is possible to avoid the feedback for n_{k-1} using an optimal quantum-control algorithm [43], although with the device used here, it leads to longer questioning time and thus degraded counting fidelities.

VI. SINGLE-SHOT PHOTOCOUNTING

We now demonstrate the number-resolved photocounting using questions Q_1 and Q_2. The device thus resolves photon numbers from 0 to 3. The feedback of n_1 is performed with minimal added latency (200 ns) using Quantum Machines' FPGA-based control system (OPX). To benchmark the photocounter, we send at its input a sech waveform in a coherent state of complex amplitude α using a microwave source [Fig. 5(a)]. This state is caught in the memory using an optimal pump followed by the two binary questions Q_1 and Q_2 that reveal a number $n_2 = [u_2 u_1]_2$ between 0 and 3. Owing to an active reset, the counter presents a short nondeterministic average dead time of 4.5 μs. The memory is reset by applying a release pump on the JRM that empties its photons into the transmission line. The qubit is reset to its ground state using a measurement-based feedback loop.

In an ideal photocounter, the distribution of n_2 follows a Poisson distribution modulo 4, $P_{n_2} = e^{-|\alpha|^2} \sum_j [|\alpha|^{2(n_2+4j)}/(n_2 + 4j)!]$ [dashed lines in Fig. 5(b)]. The measured probabilities P_{n_2} (green diamonds) qualitatively follow the ideal Poisson distribution. However, we obtain a more quantitative agreement by solving a master equation that takes into account imperfections like the finite lifetimes of the memory and qubit, the nonzero effective temperature, and nonlinear terms [44] $\hat{\mathcal{H}}_K = -K\hat{m}^{\dagger 2}\hat{m}^2 - K_e|e\rangle\langle e|\hat{m}^{\dagger 2}\hat{m}^2$. In the following, owing to large but uncertain value of the catch efficiency η, we set it to 1 in the simulations. The transmon qubit nonlinearity induces a self-Kerr term on the memory with rate $K/2\pi = 27$ kHz. When the transmon is excited in $|e\rangle$, the self-Kerr rate is offset by $K_e/2\pi = 75$ kHz. All the above parameters are calibrated using independent measurements (Appendix B).

A more stringent test for this model consists in predicting the measurement backaction on the quantum state of the incoming mode. Using the qubit, it is possible to perform a Wigner tomography of the collapsed quantum state of the memory conditioned on the outcome n_2 of the counter [45–47] (Appendix G). The top panels of Fig. 5(c) show the Wigner functions for n_2 from 0 to 3 after catching a coherent state of amplitude $|\alpha| = \sqrt{0.5}$. The bottom panels show the computed Wigner functions using our model above. For an outcome n_2, an ideal photocounter would project the incoming state $|\psi\rangle$ into $|\psi_{n_2}\rangle \propto \sum_j |n_2 + 4j\rangle\langle n_2 + 4j\,|\psi\rangle$. Given the small mean photon number $|\alpha|^2 = 0.5$, the ideal state is close to Fock state $|n_2\rangle$. The measured Wigner functions $W(\beta)$ for $n_2 \leq 2$ are indeed close to what would be obtained for pure Fock states $|n_2\rangle$. However, for $n_2 = 3$, the relaxation of both memory and qubit induce a mixture of various Fock states, and the Wigner function does not exhibit the expected fringes. To quantify this agreement, we compute the fidelity $\mathcal{F}(\rho, \rho_{n_2})$ between the collapsed quantum state of the memory ρ and the ideal projected quantum state $\rho_{n_2} = |\psi_{n_2}\rangle\langle\psi_{n_2}|$. Many definitions of fidelity exist for mixed states. We chose the fidelity [48,49] $\mathcal{F}(\rho, \rho') = \mathrm{Tr}(\rho\rho') + \sqrt{[1 - \mathrm{Tr}(\rho^2)][1 - \mathrm{Tr}(\rho'^2)]}$, which can be computed in

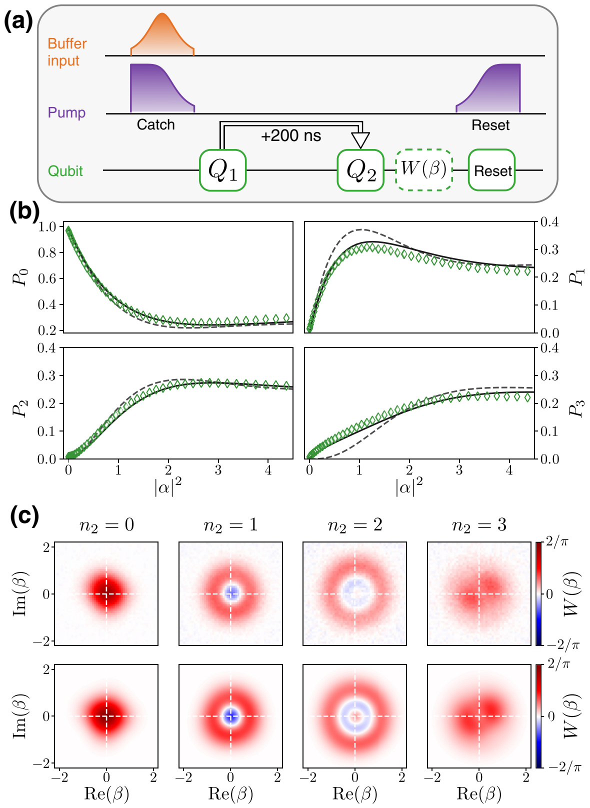

FIG. 5. Photocounting coherent states. (a) Pulse sequence showing an incoming mode on the buffer with a coherent state of amplitude α and the optimal shape of the pump to catch the wavepacket with minimal distortion (Appendix D). The qubit performs photocounting bit by bit with pulse sequences Q_k's described in Fig. 4. Q_2 uses the outcome of Q_1 in a feedback protocol that adds as little as 200-ns delay. Finally, a direct Wigner tomography of the memory can be performed [45–47] before the memory and qubit are reset. (b) Green diamonds: measured probabilities P_{n_2} of finding a number $n_2 = n \bmod 4$ photons as a function of the mean photon number $|\alpha|^2$ of the incoming coherent state after 200 000 runs of the sequence. Dashed lines: modulo-4 Poisson distribution. Solid lines: master-equation solution without any free parameter. (c) Corresponding measured (top) and simulated (bottom) Wigner functions for $\alpha = \sqrt{0.5}$ mean photons (Appendices G and F). From left to right, the Wigner function is heralded on the counter outcome $n_2 = 0$, 1, 2, and 3 out of a total of 44 000 realizations per pixel.

a numerically robust manner from the measured Wigner functions since $\mathrm{Tr}(\rho_1\rho_2) = \pi \int W_{\rho_1}(\beta) W_{\rho_2}(\beta) d^2\beta$. From the measured Wigner functions in Fig. 5(c) (top panels), we obtain fidelities of 86, 52, 32, and 4.9% for $n_2 = 0$, 1, 2, and 3, respectively. This deviation from the ideal case is well captured by our model, which predicts the measured collapsed quantum states with fidelities between top and bottom panels of Fig. 5(c) above 97% for the four outcomes of the counter. Simulations show that the dominant origin for the nonidealities is the qubit and memory relaxation (Appendix H).

TABLE I. Probabilities of getting the outcome m if the incoming mode is in Fock state $|n\rangle$. The probabilities are computed using the master equation validated by Fig. 5. The uncertainties correspond to the range of possible values on the catch efficiency η. Diagonal terms are all above 25%, which would correspond to a completely random counter with four possible outcomes.

$\mathcal{P}_{\lvert n\rangle}(m)$	$\lvert 0\rangle$	$\lvert 1\rangle$	$\lvert 2\rangle$	$\lvert 3\rangle$
$m = 0$	**99%**	$(7 \mp 4)\%$	$(24 \mp 3)\%$	$(9 \mp 4)\%$
$m = 1$	<1%	$\mathbf{(76 \pm 3)\%}$	$(4.2 \pm 0.2)\%$	$(27 \pm 1)\%$
$m = 2$	<1%	$(1.03 \pm 0.01)\%$	$\mathbf{(71 \pm 3)\%}$	$(9.7 \pm 0.3)\%$
$m = 3$	<1%	$(16 \pm 1)\%$	$(1.5 \pm 0.1)\%$	$\mathbf{(54 \pm 2)\%}$

Since the model is backed up by the photon-number statistics and by the Wigner tomography, we can compute the probabilities $\mathcal{P}_{|n\rangle}(m)$ that the counter would have measured m mod 4 if a Fock state $|n\rangle$ had been sent at the input (see Table I). If the detector is giving totally random outcomes, the probabilities are equal to 25% since there are four possible answers. Here, we obtain fidelities $\mathcal{P}_{|n\rangle}(n)$ well above 25% and infidelities $\mathcal{P}_{|n\rangle}(m \neq n)$ smaller or of the same order of 25%. Interestingly, downgraded to a photodetector that clicks when $m \neq 0$, these figures imply a detection fidelity of $1 - \mathcal{P}_{|1\rangle}(0) = 93 \pm 4\%$ for a single photon. The model reveals three main sources of errors: the finite lifetimes of the memory and qubit, and the rate K_e (Appendix H). The finite qubit lifetime affects the various n_2 values differently owing to the choice of encoding in the qubit state during questions Q_k's. It is possible to choose which photon number to affect the least by swapping the roles of $|g\rangle$ and $|e\rangle$. The photon number corresponding to the qubit being in the excited state after each question is the one with maximum error. Here, we choose to minimize the error on $n_2 = 0$ and thus minimize the dark count of the counter to a measured probability of 3% [measurement of $1 - P_0$ at $\alpha = 0$ in Fig. 5(b)]. When the incoming photon number increases, the memory relaxation starts to limit the fidelity since the loss rate of the memory increases with photon number. It explains most of the decrease of fidelity with photon number from 99% down to $54 \pm 2\%$. Finally, because of the nonzero K_e, during the time T_k of the question Q_k, the qubit acquires an additional parasitic phase that rapidly increases with the photon number resulting in larger infidelities for higher n.

VII. CONCLUSION

We develop a photocounter using measurement-based feedback that is able to resolve the photon number from $n = 0$ up to $n = 3$ in a propagating microwave mode. The counter features a time-resolved power meter able to determine the envelope of the incoming waveform *in situ*, which optimizes the detection efficiency up to $\eta = 0.96 \pm 0.04$. Future devices with longer lifetimes would considerably improve the fidelities $\mathcal{F}$ above. The reset would then release a faithfully collapsed quantum state into the line, making the photocounter quantum nondemolition. The counter would then quickly scale up to resolve higher photon number thanks to its logarithmic complexity. The photocounter can also be used in a degraded mode to measure parity by asking a single question Q_1 as in Refs. [20,21], and thus perform propagating Wigner tomography [50]. Microwave photodetection and photocounters enable quantum-optics-like experiments in the microwave range and facilitate the implementation of a quantum network. For instance, photodetection has made possible the entanglement between remote stationary qubits [22,34,35]. However, any protocol requiring feedback on the photon number in a propagating mode needs a single-shot photocounter. Therefore, a direct application consists in reaching the quantum limit for the discrimination between two coherent states [51], with obvious applications in quantum sensing.

ACKNOWLEDGMENTS

We are grateful to Olivier Buisson, Michel Devoret, Zaki Leghtas, Danijela Marković, Mazyar Mirrahimi, Alain Sarlette for discussions. We acknowledge IARPA and Lincoln Labs for providing a Josephson Traveling-Wave Parametric Amplifier. The device is fabricated in the cleanrooms of Collège de France, ENS Paris, CEA Saclay, and Observatoire de Paris. The feedback code is developed in collaboration with Quantum Machines [52]. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 820505.

APPENDIX A: MEASUREMENT SETUP

The sample and its fabrication are described in Ref. [28]. The sample is cooled down to 24 mK in a BlueFors LD250 dilution refrigerator. The diagram of the microwave wiring is given in Fig. 6. The buffer, memory, qubit, and readout pulses are generated by modulation of continuous microwave tones produced, respectively, by generators E8752D from Keysight, SGS100A from Rohde&Schwarz, SGS100A from Rohde&Schwarz, and SynthHD PRO from Windfreak set, respectively, at frequencies $f_b + 50$, $f_m - 120$, $f_q + 200$, and $f_r + 51$ MHz. The pump pulses are also generated by modulation of continuous microwave tone, however, the local oscillator at $f_b - f_m + 170$ MHz is produced by mixing the buffer and the memory rf sources for phase stability. The readout is modulated through a single sideband mixer while the others are modulated via IQ mixers. The IF modulation pulses are generated by nine channels of an OPX from Quantum Machines with a sample rate of 1 GS/s. The acquisition is performed, after down-conversion by their local oscillators, by digitizing a 51-MHz (readout) or a 50-MHz (buffer) signal with the 1 GS/s analog-to-digital converter (ADC) of the OPX from Quantum Machines. The signals coming out of the buffer mode and of the readout mode are multiplexed into a single transmission line using a diplexer before getting amplified by a traveling-wave parametric amplifier [53] (TWPA, provided by IARPA and the Lincoln Labs). The TWPA is pumped at a frequency $f_{\mathrm{TWPA}} = 7.636$ GHz and at a power that allowed the TWPA to reach a system efficiency of 18% from the buffer output to the ADC. The signal coming out of the buffer mode is filtered using a 20-cm waveguide WR62 with a cutoff frequency at 9.8 GHz in order to prevent the strong pump of the JRM from reaching the TWPA and reciprocally. The next stage of amplification is performed by a HEMT amplifier (from Caltech) at 4 K and by two room-temperature amplifiers.

APPENDIX B: SYSTEM CHARACTERIZATION AND FLUX DEPENDENCE

Using a vector network analyzer we measure the buffer resonance frequency as a function of the current running through a superconducting coil directly above the sample. The extracted buffer frequency ω_b is displayed in Fig. 7(a). The current is generated by applying a voltage V_{coil} to a resistor in series with the coil. The periodicity of the buffer frequency allows us to convert the voltage V_{coil} into a flux Φ_{ext} through the four inner loops of the JRM.

Even though the qubit consists in a single junction transmon, its frequency ω_q has a slight flux dependence due to its coupling with the memory. The qubit frequency, as a function of the flux, is extracted from Ramsey oscillations [Fig. 7(c)]. With these measurements, we are also able to extract the qubit coherence time T_2 as a function of flux Φ_{ext} [solid line in Fig. 7(d)].

The memory cannot be probed directly in reflection nor in transmission with the measurement setup. To measure its frequency ω_m [Fig. 7(b)], we use the qubit to determine at what excitation frequency the memory gets populated. We send a probe pulse on the memory via its weakly coupled port followed by a conditional π pulse on the qubit at ω_q. The qubit is thus excited only if the memory has zero photons. Measuring the qubit average excitation as a function of probe frequency leads to determining the frequency ω_m at which the state $|0\rangle$ is most depleted. We also measure the relaxation times of the qubit $T_{1,q}$ [see Fig. 7(d)]. The qubit decoherence time is limited by the relaxation since T_2 is close to $2T_{1,q}$.

We extract the buffer self-Kerr rate K_{bb} from the dependence of its frequency ω_b as a function of probe power [Fig. 8(a)]. To measure the pump-buffer cross-Kerr rate K_{bp} [Fig. 8(b)], we measure ω_b while driving the pump at various powers. The pump is driven off resonance from $\omega_b - \omega_m$ to avoid frequency conversion. The buffer self-Kerr and buffer-pump cross-Kerr rates both vanish at the same flux point [31], which we hence choose as our working point. A nonzero cross-Kerr rate would indeed make

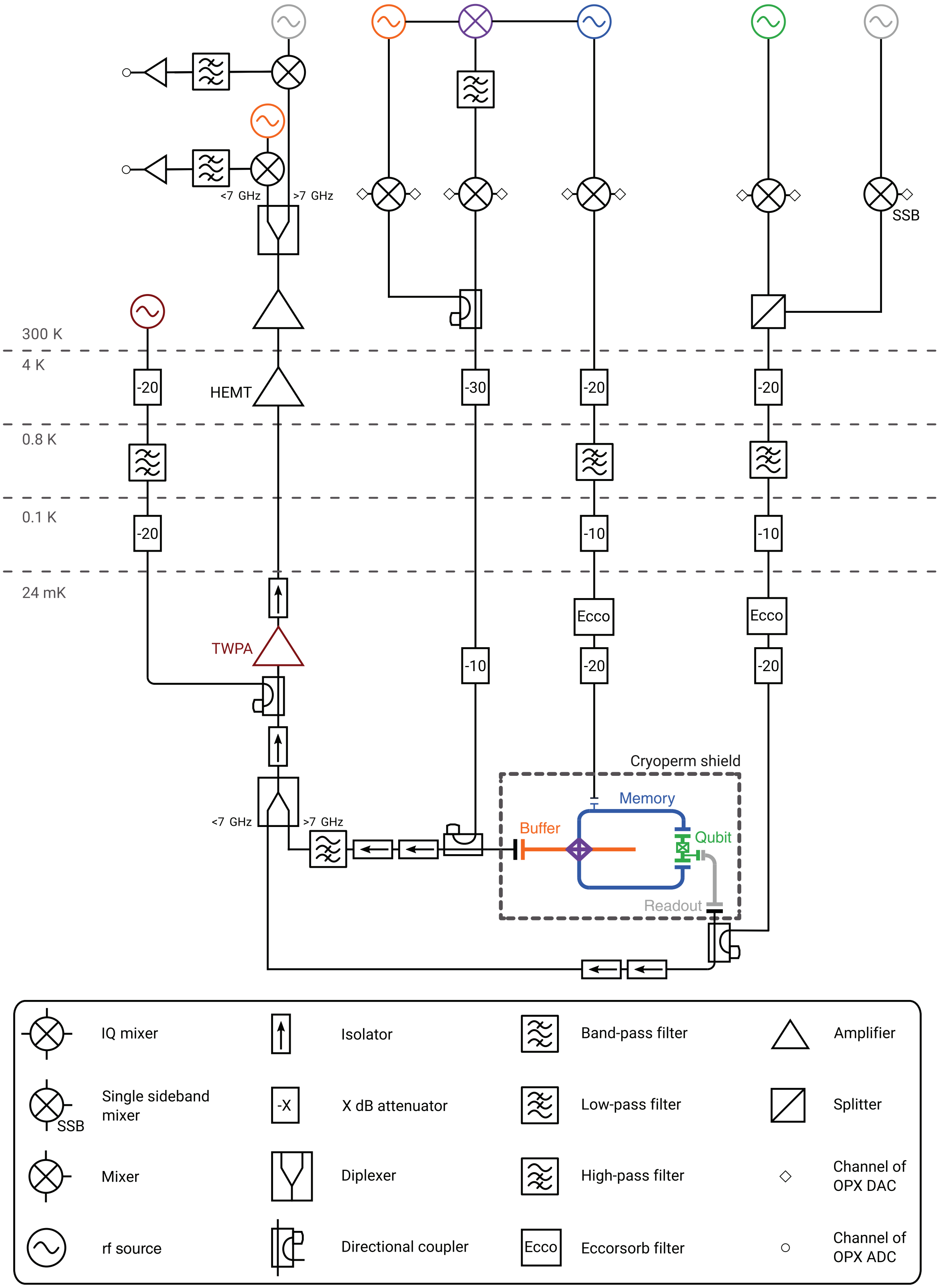

FIG. 6. Scheme of the measurement setup. The rf sources color refers to the frequency of the matching element in the device up to a modulation frequency. Identically colored sources represent a single instrument with a split output.

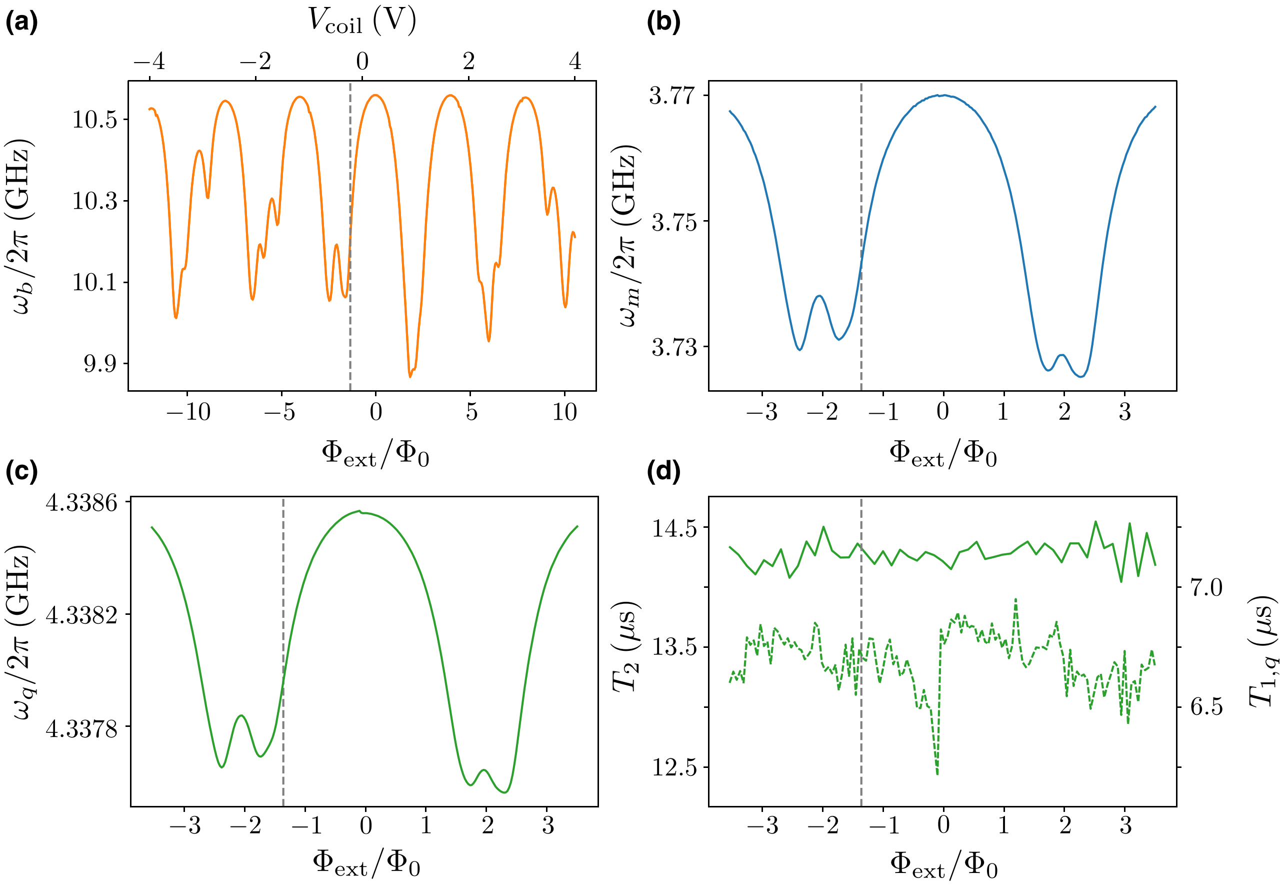

FIG. 7. (a) Buffer frequency, (b) memory frequency, (c) qubit frequency, (d) qubit decoherence time T_2 (dashed line), and lifetime $T_{1,\rm q}$ (solid line) as a function of flux $\Phi_{\rm ext}$ in the inner loops of the JRM. Notice that the flux range is different in (a) compared to (b)–(d). Vertical dashed line: working point for the main text.

the pump optimization more challenging for catch and reset operations.

The measurement of the memory self-Kerr rate K and the qubit-dependent nonlinear rate K_e are done in a previous cool down by monitoring the average phase acquired by a coherent state in the memory mode as a function of time while varying the mean photon number and the initial qubit state. Having prepared the qubit in either $|g\rangle$ or $|e\rangle$, we load the memory with a coherent state of amplitude $\alpha = \sqrt{n}$. We then wait for a time $t_{\rm int}$. Finally, we release the state of the memory into the transmission line and record the average phase $\phi(t_{\rm int})$ of the released pulse. The detuning $\delta\omega_m$ between the resonant frequency of the memory ω_m and a reference resonant frequency (when the memory is in the vacuum state and the qubit in $|g\rangle$) can be determined as $\delta\omega_m = d\phi/dt_{int}$. The slope of $\delta\omega_m$ as a function of mean photon number n then gives the self-Kerr rate K ($K_e + K$) when the qubit is prepared in $|g\rangle$ ($|e\rangle$). The rates K and K_e are plotted as a function of flux in Fig. 8(c).

Using a populated Ramsey protocol [see details in Fig. 10] as a function of flux, we also extract the qubit-memory dispersive coupling χ [Fig. 8(d)]. It is also performed in a previous cool down.

APPENDIX C: READOUT OPTIMIZATION

The readout strategy is a compromise between readout speed, fidelity, and QNDness. Note that the feedback protocol of the photocounter requires a QND measurement so that non-QNDness limits the counter fidelity. In order to make fast and faithful qubit measurements, we implement a CLEAR-like sequence [39] with amplitude $r_{\rm in}(t)$ shown in Fig. 9(a). The QNDness of the readout is limited by the possible ionization of the transmon out of the qubit subspace [54,55]. We find that not only this constraint limits the amplitude of the readout pulse but also that the ionizaiton probability increases with the occupation of the memory mode [Fig. 9(b)]. In future design, the efficiency of the photocounter could be improved by using less sensitive coupling schemes [56–58].

In order to determine the state of the qubit as a function of the reflected signal with the best fidelity, we use a set of

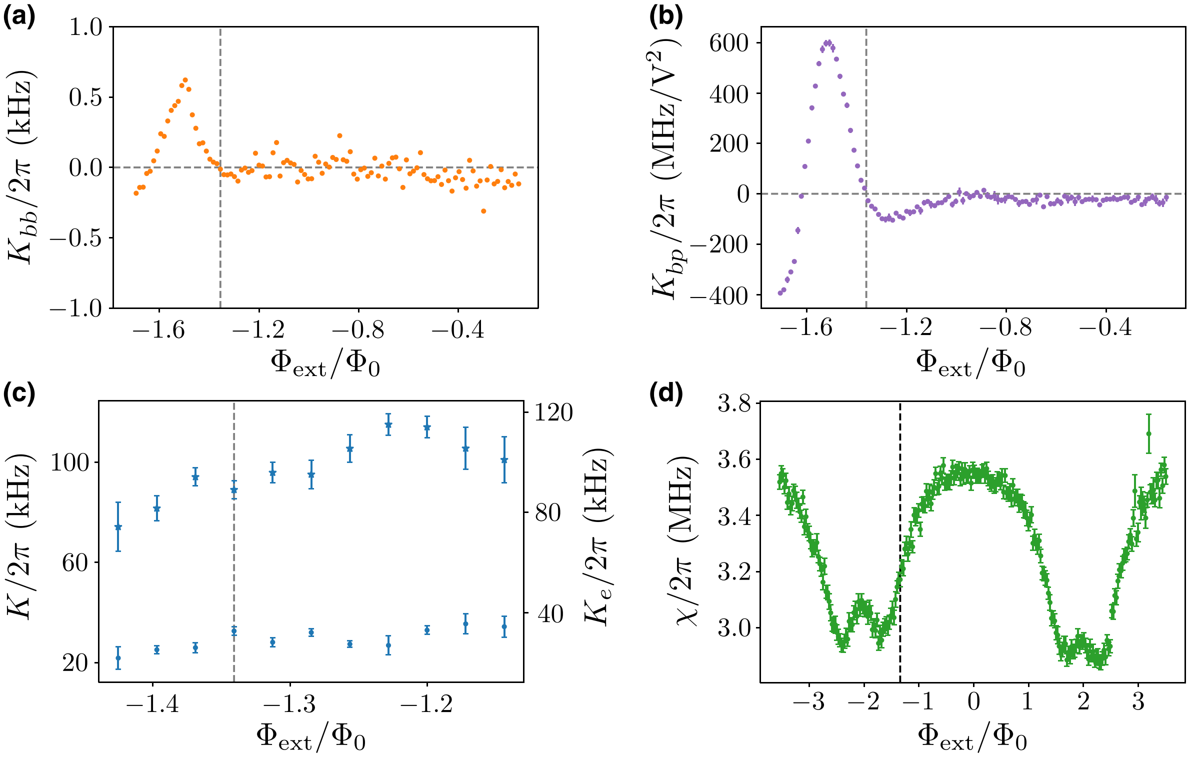

FIG. 8. Rates of nonlinear terms in the device as a function of the external flux $\Phi_{\rm ext}$. Notice that the flux range is different for each panel. (a) Buffer self-Kerr rate K_{bb}. (b) Pump-buffer cross-Kerr rate K_{bp}. (c) Dots, memory self-Kerr rate K; stars, nonlinear rate K_e. (d) Dispersive shift χ between qubit and memory.

optimized demodulation weights that we compute to maximize the complex signal difference between the ground and excited states as shown in Ref. [59]. It is convenient to quantify the readout error using the overlap ϵ_0 between the two Gaussian distributions corresponding to the two qubit states [60].

The qubit temperature is measured by repeatedly measuring the qubit, recording the demodulated signal from the readout into a complex histogram [such as the one shown in Fig. 9(b)] and fitting it with a set of two two-dimensional (2D) Gaussians of equal width. The temperature is then extracted by taking the ratio of the amplitudes of the two Gaussians. For additional precision, the center of the Gaussian corresponding to the qubit being in the excited state $|e\rangle$ is estimated by doing the same measurements after performing a π pulse such that the final fit only had two free parameters: the center of the Gaussian corresponding to $|g\rangle$ and the qubit

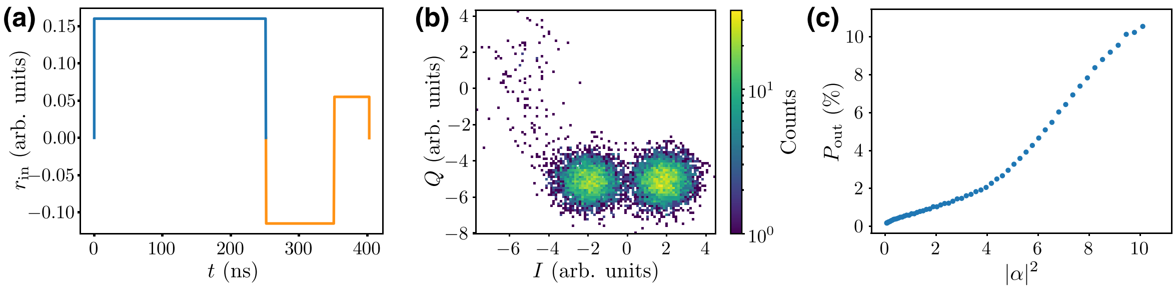

FIG. 9. Readout optimization. (a) CLEAR-like readout pulse sequence. Driving amplitude $r_{\rm in}$ of the readout as a function of time t. Blue, readout excitation; orange, readout reset. (b) Histogram of the two demodulated quadratures I and Q of the reflected readout pulse for 10^4 realizations after applying a $\pi/2$ pulse on the qubit. The two peaks correspond to the $|g\rangle$ and $|e\rangle$ states of the qubit. The few points in the upper-left corner correspond to the transmon in an ionized state. (c) Probability to observe the transmon outside of its qubit subspace as a function of the mean number of photons inside of the memory for the readout power used in the main text.

temperature. We find an effective temperature of 33 mK.

APPENDIX D: OPTIMAL CATCHING PUMP

In this section, we derive the optimal pump to catch an arbitrary wavepacket with a bandwidth smaller than the bandwidth of the buffer $\kappa_b = 2\pi \cdot 20$ MHz. We first derive the optimal pump to catch an incoming wavepacket assuming $\kappa_m = 0$ and we then show that a small memory relaxation rate κ_m and a cross-Kerr rate K_{bp} do not prevent the catch from being complete.

1. Ideal case

Let us consider the Langevin equations for the buffer b and memory m with a conversion pump p in the frame rotating with b_{in} and m

$$\frac{db}{dt} = -\frac{\kappa_b}{2}b(t) - g_3 p^*(t)m(t) + \sqrt{\kappa_b}b_{\mathrm{in}}(t)$$
$$\frac{dm}{dt} = g_3 p(t)b(t),$$

where, for simplicity, we assume that the external flux used is chosen such that all the self-Kerr and cross-Kerr terms cancel out. Note that an arbitrary choice of phase reference allows us to constrain b to be a real function.

We start by parametrizing the equations with dimensionless variables using $\tau = (\kappa_b/2)t$, $u = 2g_3 p/\kappa_b$

$$\dot{b} = -b - u^* m + \frac{2}{\sqrt{\kappa_b}}b_{\mathrm{in}},$$
$$\dot{m} = ub,$$

where the dots denote the derivatives with respect to τ.

Catching the incoming wavepacket b_{in} perfectly comes down to finding the pump $u(\tau)$ such that $b_{\mathrm{out}} = 0$ uniformly. Since $b_{\mathrm{in}} + b_{\mathrm{out}} = \sqrt{\kappa_b}b$, u is the solution of the following differential equations:

$$um^* = b - \dot{b}, \tag{D1}$$

$$\dot{m} = ub. \tag{D2}$$

For any signal with a bandwidth lower than the buffer coupling rate κ_b, these equations can be solved numerically. In the following subsection, we focus on the case of a sech input waveform, where the calculation can be carried out analytically.

a. Case of an incoming hyperbolic secant waveform

In the experiment, we frequently use an incoming hyperbolic secant waveform $b(\tau) = [\sqrt{(\lambda/2)}/2]\mathrm{sech}(\lambda\tau/2)$. To do so, we remark that

$$y = |m|^2 + b^2$$

is a flat output [61], meaning that m, u, and b can be expressed as functions of $y, \dot{y}$, and $\ddot{y}$. Combining Eq. (D1) and (D2), we get $m^*\dot{m} = (b - \dot{b})b$. Taking the real part and using the limited bandwidth ($\dot{y} \leq 2y$) and the assumption that there is no loss ($0 \leq \dot{y}$), we get

$$b^2 = \dot{y}/2, \quad |m|^2 = y - \dot{y}/2. \tag{D3}$$

Setting $y = 1/(1 + e^{-\lambda\tau})$ with $0 \leq \lambda \leq 2$, using Eq. (D3), we get $b(\tau) = [\sqrt{(\lambda/2)}/2]\mathrm{sech}(\lambda\tau/2)$ as desired and $|m| = [\sqrt{(e^{\lambda\tau} + 1 - \lambda/2)}/2]\mathrm{sech}(\lambda\tau/2)$. Multiplying Eq. (D1) by its complex conjugate, we obtain $|u| = (b - \dot{b})/|m|$. From Eq. (D1), we can also see that $\arg(u) = \arg(m)$. Hence, there is a function θ such that $m = |m|e^{i\theta}$ and $u = |u|e^{i\theta}$. By multiplying Eq. (D2) by m^* and using Eq. (D1) one gets $\dot{m}m^* = (b - \dot{b})b$. Since b is real, the imaginary part, yields $\dot{\theta} = 0$. For simplicity, we choose $\theta(\tau) = 0$, which leads to

$$u = \frac{b - \dot{b}}{|m|}. \tag{D4}$$

Finally, we find

$$u(\tau) = \sqrt{\frac{\lambda/2}{e^{\lambda\tau} + 1 - \lambda/2}}\left[1 + \frac{\lambda}{2}\tanh(\lambda\tau/2)\right]. \tag{D5}$$

Going back to the original time variable t, we conclude that an incoming wavepacket with a shape $b_{\mathrm{in}}(t) = \sqrt{\lambda/8\kappa_{\mathrm{b}}}\mathrm{sech}(\lambda\kappa_{\mathrm{b}}t/4)$ is perfectly caught by a pump $p_{\mathrm{opt}}(t) = (2g_3/\kappa_{\mathrm{b}})\sqrt{(\lambda/2)/(e^{\lambda\kappa_{\mathrm{b}}t/2} + 1 - \lambda/2)}[1 + (\lambda/2)\tanh(\lambda\kappa_{\mathrm{b}}t/4)]$.

2. Finite memory lifetime

In order to account for the memory relaxation rate κ_m, the Langevin equations become

$$\frac{db}{dt} = -\frac{\kappa_b}{2}b(t) - g_3 p^*(t)m + \sqrt{\kappa_b}b_{\mathrm{in}}(t)$$
$$\frac{dm}{dt} = -\frac{\kappa_m}{2}m(t) + g_3 p(t)b(t).$$

Without loss of generality, we assume that b_{in} and p are real, hence m and b are also real. Using the same definition for y and introducing $\varepsilon = \kappa_m/\kappa_b$, we get the following modified version of Eq. (D3) to derive b and m as algebraic functions of y and $\dot{y}$.

$$(1 + \varepsilon)b^2 = \dot{y}/2 + \varepsilon y, \quad (1 + \varepsilon)|m|^2 = y - \dot{y}/2. \tag{D6}$$

Given Eq. (D4), u can be expressed as an algebraic function of y, $\dot{y}$, and $\ddot{y}$. In this case the no-loss assumption is replaced by the weaker constraint that the ratio between the outgoing power $-dy/dt$ and the total energy y is smaller than κ_m, i.e., $dy/dt \geq -\kappa_m y$ (i.e., $\dot{y}/2 + \varepsilon y \geq 0$). The bandwidth limit $dy/dt \leq \kappa_b y$ remains valid (i.e., $\dot{y} \leq 2y$).

To carry on the calculation analytically, we set $y = 1/(e^{2\varepsilon\tau} + e^{-\lambda\tau})$ so that

$$b(\tau) = \sqrt{\frac{\lambda/2+\varepsilon}{1+\varepsilon}} \frac{1}{e^{(\lambda/2+2\varepsilon)\tau} + e^{-\lambda\tau/2}}.$$

We also get

$$m(\tau) = \sqrt{e^{(\lambda+2\varepsilon)\tau} + \frac{1-\lambda/2}{1+\varepsilon}} \frac{1}{e^{(\lambda/2+2\varepsilon)\tau} + e^{-\lambda\tau/2}}.$$

From the above expressions for b and m, we can then compute u using Eq. (D4). Given the small value of $\varepsilon \approx 0.002$ in the device of the main text, we choose to neglect the memory relaxation and to use the results from the ideal case above.

3. Finite cross-Kerr rate

Even in the presence of a small cross-Kerr rate K_{bp} between the buffer and the pump, an optimal catch pump can be found, which guarantees that no signal is reflected, i.e., $b_{\text{out}} = 0$. The modified Langevin equations are as follows:

$$\frac{db}{dt} = -\left[\frac{\kappa_b}{2} + iK_{bp}|p(t)|^2\right] b(t) - g_3 p^*(t) m(t) + \sqrt{\kappa_b} b_{\text{in}}(t),$$

$$\frac{dm}{dt} = -\frac{\kappa_m}{2} m(t) + g_3 p(t) b(t).$$

Introducing the dimensionless cross-Kerr rate $k = K_{bp}\kappa_b/g_3^2$, we get a modified version of Eqs. (D1) and (D2)

$$um^* = b - \dot{b} + ik|u|^2 b,$$

$$\dot{m} = -\varepsilon m + ub.$$

Since b is real, the real quantity $y = |m|^2 + b^2$ can still be used to parametrize the system, despite the fact that m and u are now complex. The values of b and $|m|$ can still be expressed as functions of y and $\dot{y}$ by Eq. (D6). The modulus $|u|$ of the pump is obtained by solving

$$|u|^2|m|^2 = (b - \dot{b})^2 + k^2|u|^4 b^2.$$

The argument θ_m of m results from the integration

$$\theta_m(\tau) = \theta_m(0) + k\int_0^\tau \frac{|u(s)|^2 b(s)^2}{|m(s)|^2} ds,$$

where $|u|$, $|m|$, and b are algebraic functions of y, $\dot{y}$, and $\ddot{y}$. The argument θ_u of u is given by the argument of $m(b - \dot{b} + ik|u|^2 b)$, which coincides then with the argument of $(\dot{m} + \varepsilon m)/b$.

Using the above derivation, one sees that finding the optimal pump in the case of cross-Kerr effect requires not only to adjust the envelope of the pump, as done in the main text, but also adjusting the phase of the pump θ_u dynamically to compensate for the time-dependent buffer frequency shift.

APPENDIX E: DIFFERENT METHODS FOR MEASURING THE MEAN PHOTON NUMBER

We use several methods to measure the mean photon number $\langle n\rangle$ in the memory in order to calibrate the buffer and memory displacement pulses (Fig. 10). The experiment begins by a displacement pulse on the memory mode with a driving voltage $\mu\alpha$, where μ is a conversion factor between voltages and amplitudes to be determined. The following procedures then determine the mean photon number $\langle n\rangle = |\alpha|^2 + n_{\text{th}}$ as a function of the driving voltage by different ways and thus calibrate μ. n_{th} is the residual equilibrium thermal photon number in the memory.

1. Photon number selective π pulse

The first method relies on the possibility to perform a π pulse $\Pi_{|n\rangle}$ conditionally on the photon number n. It is done by driving the qubit at frequency $\omega_q - \chi n$ with a long enough pulse so that the frequency spreading is smaller than $\chi/2$. The pulse maps the probability to have n photons $P(n, \alpha)$ into the measured probabilities $P_{|n\rangle,\alpha}(e)$ for the qubit to be found in its excited state [Fig. 10(a)]. Fitting the distribution $P(n, \alpha)$ for each α by a Poisson distribution, we calibrate μ neglecting the thermal population. A limitation of this method occurs at high photon number. Indeed, the dispersive shift χ slightly depends on photon number n, so that the qubit drive frequency is off resonant.

2. Vacuum detector

To calibrate the conversion factor μ at high photon numbers $|\alpha|^2 \gg 1$, we perform another method, which is to use the qubit as a vacuum detector [28]. Applying a π pulse $\Pi_{|0\rangle}$ encodes the probability that the memory is empty into the probability for the qubit to be in the excited state. Now, after a waiting time t, the memory has relaxed and, neglecting n_{th} for the large $|\alpha|^2$, the measured probability $P_{|0\rangle,\alpha}(e)$ evolves following $\exp(-|\alpha|^2 e^{-t/T_{1,\text{m}}})$ [Fig. 10(b)]. Fitting the value of μ for each value of $\mu\alpha$ to match this expression with the measured $P_{|0\rangle,\alpha}(e, t)$ leads to an accurate determination of the conversion factor μ as a function of α. This photon number calibration has a higher range than the previous one but is less sensitive for low average photon numbers.

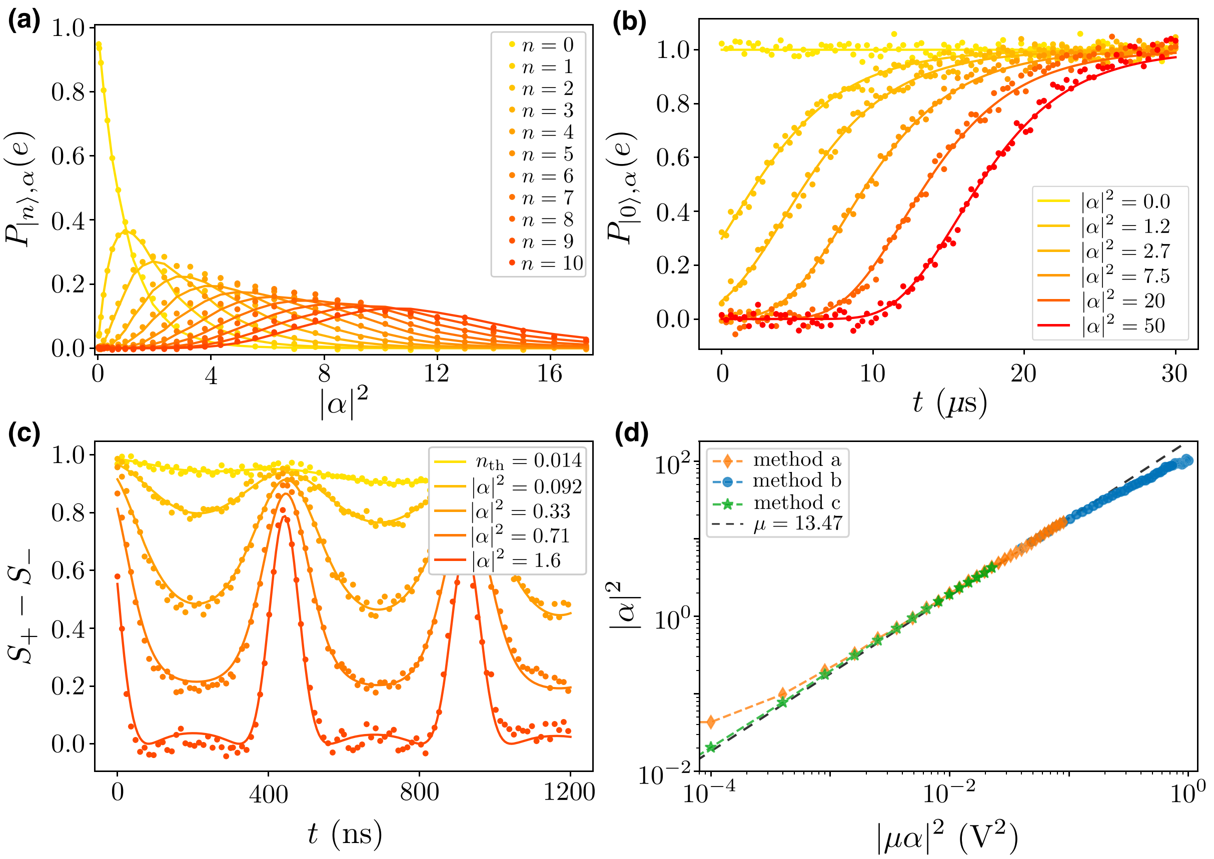

FIG. 10. Three methods for calibrating the memory-displacement amplitude. The measurements are performed on a previous cool down. (a) Photon-number selective π pulse. Dots: measured probability to have n photons in the memory as a function of $|\alpha|^2$. Solid lines: Poisson distribution fitted to calibrate the mean photon number on the x axis. (b) Vacuum detector. Dots: probability $P_{|0\rangle,\alpha}(e)$ that the memory is empty as a function of waiting time for various preparation amplitudes. Solid lines: fit of the measured probabilities using the expression for memory relaxation in the text. (c) Populated Ramsey. Dots: signal difference $S_+ - S_-$ between two encodings of the Ramsey-like interferences in the presence of various mean photon numbers $\langle n\rangle$. Solid lines: theoretical prediction allowing to calibrate the displacement amplitude and the thermal occupancy n_{th}. (d) Result of the calibration using the three methods: photon number selective π pulse in orange diamonds, vacuum detector in blue dots and populated Ramsey in green stars. The black dashed line represents the overall fitted value for μ.

3. Populated Ramsey oscillations

Our last method to calibrate the conversion factor μ relies on a Ramsey-like sequence [62] [Fig. 10(c)]. After the coherent displacement of the memory, we prepare the qubit in an equal superposition of ground and excited states by applying an unconditional $\pi/2$ pulse. After a waiting time t, the phase of the superposition increases by χnt for each Fock state $|n\rangle$. We then apply a second unconditional $\pm\pi/2$ pulse giving the signal $S_\pm$. The signal difference is given by $S_+ - S_- = \cos[\langle n\rangle \sin(\chi t)] \exp\{\langle n\rangle[\cos(\chi t) - 1] - t/T_2\}$ from which we extract the mean photon number $\langle n\rangle$. Without driving the memory, the measured mean number gives the thermal population of the memory $n_{\mathrm{th}} = 0.014$ corresponding to an effective temperature of 44 mK. Offsetting the measured $\langle n\rangle$ by this thermal occupation leads to a calibration of μ. This last method has a good sensitivity at low photon numbers, however, it cannot be used for large photon numbers where the pattern becomes insensitive to $\langle n\rangle$.

4. Comparison

In Fig. 10(d), we show the outcome of the three methods by plotting the measured $|\alpha|^2$ as a function of driving power. The methods agree over their respective ranges. For large mean photon number $|\alpha|^2 > 20$, due to memory self-Kerr, the mean photon number is expected to differ and be smaller than the linear behavior $|\mu\alpha|^2$.

APPENDIX F: NUMERICAL MODEL

We simulate our system using the QuantumOptics.jl library [63].

The device Hamiltonian reads [28]

$$\begin{aligned}\hat{\mathcal{H}}/\hbar = {} & \omega_b\hat{b}^\dagger\hat{b} + \omega_m\hat{m}^\dagger\hat{m} + \frac{\omega_q}{2}\hat{\sigma}_z \\ & + g_3 p\hat{m}^\dagger\hat{b} + g_3^* p^*\hat{m}\hat{b}^\dagger \\ & - \chi\hat{m}^\dagger\hat{m}|e\rangle\langle e| - K\hat{m}^{\dagger 2}\hat{m}^2 - K_e|e\rangle\langle e|\hat{m}^{\dagger 2}\hat{m}^2.\end{aligned}$$

To simplify the model, we restrict the transmon to its first two levels and we do not consider the readout resonator and its dispersive coupling to the qubit. We simulate the readout of the qubit by an instantaneous projective measurement taking place at half of our experimental readout duration. During the readout time, before and after the projection, the system evolves freely. We also take into account the overlap error ε_o [60] in the readout, which we measure to be below 1%. Moreover, we consider the catch of the wavepacket incoming onto the buffer to be optimal (Appendix D). Thus, we further reduce the numerical Hilbert space by putting aside the buffer and the pump. The catch is then simulated by an instantaneous displacement on the memory field.

Finally, we model our system in the memory and qubit rotating frame using the following Hamiltonian:

$$\begin{aligned}\hat{\mathcal{H}}/\hbar = {} & -\chi\hat{m}^\dagger\hat{m}|e\rangle\langle e| - K\hat{m}^{\dagger 2}\hat{m}^2 - K_e|e\rangle\langle e|\hat{m}^{\dagger 2}\hat{m}^2 \\ & + \mathrm{Re}[f(t)]\hat{\sigma}_x + \mathrm{Im}[f(t)]\hat{\sigma}_y \qquad \text{(F1)}\end{aligned}$$

with $f(t)$ the complex envelope containing all the qubit drives. Using a time-dependent Hamiltonian allows us to simulate the optimal counting with the questions Q_0 and Q_1. For instance, we can thus accurately take into account the finite duration of the $\pi/2$ pulses. A Lindblad master equation enables us to take into account the qubit relaxation time $T_{1,q}$ and pure dephasing time T_ϕ and the cavity lifetime $T_{1,\mathrm{m}}$ as well as temperatures of qubit and memory. We restrict the Hilbert space of the memory mode between 0 and 29 photons.

APPENDIX G: WIGNER TOMOGRAPHY

We use the method of Refs. [45–47] to directly measure the Wigner function $W(\beta) = (2/\pi)\langle\mathcal{D}_\beta\mathcal{P}\mathcal{D}_\beta^\dagger\rangle$ of the memory mode. We perform a displacement $\mathcal{D}_\beta^\dagger$ of amplitude $-\beta$ (sech-shape with $\sigma = 13$ ns) followed by a parity measurement. $\mathcal{P} = \exp(i\pi m^\dagger m)$ is the photon parity operator. The Wigner functions are measured on a 51×51 square matrix of amplitudes β where $|\mathrm{Re}(\beta)|, |\mathrm{Im}(\beta)| \leq 2.2$. The measured Wigner functions for mean photon numbers $|\alpha|^2 = 0, 1, 1.5$, and 2 are shown in Fig. 11(a). Each column corresponds to postselected measurements for a given detected photon number n_2.

Our numerical model above allows us to compute the predicted Wigner functions for each panel of the figure. The predictions are shown in Fig. 11(b). Note that these figures are obtained by computing the Wigner function directly without modeling the readout of the parity photon number after displacement.

For an arbitrary outcome n_2, the photocounter would ideally project the incoming state $|\psi\rangle$ into $|\psi_{n_2}\rangle \propto \sum_j |n_2 + 4j\rangle\langle n_2 + 4j|\psi\rangle$. We discuss nonidealities in the measurement backaction in the main text. They are mainly due to the finite lifetimes of the qubit and memory for low mean photon numbers $|\alpha|^2$.

In Fig. 11, some Wigner functions are not invariant by a phase shift as one could expect from mixtures of Fock states. These patterns in the figure indicate coherences between Fock states. Our simulations show that the coherences originate from two main phenomena. First, the photon number measurement is performed modulo 4, which preserves coherences between different photon numbers modulo 4 by projection. Second, due to the finite duration of the $\pi/2$ pulses in the pulse sequence that performs question Q_k, the encoding of the kth bit of the photon number in the qubit state is imperfect. Therefore, postselecting on the measured binary code n_2 preserves some coherence between the Fock states that compose the initial coherent state $|\alpha\rangle$. Finally, the Wigner functions appear distorted due to the memory nonlinear rates K and K_e.

The deviations from the ideal projected quantum state (fidelities in Table II) are further investigated in Appendix H.

APPENDIX H: ERROR BUDGET OF THE PHOTOCOUNTER

In this section we numerically investigate the origin of the errors on the success probabilities $\mathcal{P}_{|n\rangle}(n)$ to find n photons when the incoming wavepacket is in a Fock state $|n\rangle$ and on the QNDness, which is characterized by the fidelities $\mathcal{F}$ above. We study the error budget by sweeping one (or more) parameters independently of the others in our model.

(a) The finite qubit relaxation time $T_{1,q}$ entails different errors depending on the choice of encoding the outcome n_2 in the qubit state during questions Q_k's. This choice is done by the sign of the second $\pi/2$ pulse in the sequence of Fig. 3. For each question Q_k, the outcome on the kth bit of the photon number corresponding to the qubit excited state will get mixed with the outcome corresponding to the qubit ground state. These errors scale exponentially with $1/T_{1,q}^{\mathrm{model}}$ [Figs. 12(a) and 13(a)].

(b) The finite memory relaxation time $T_{1,m}$ causes errors except for $|n = 0\rangle$ [Figs. 12(b) and 13(b)]. The dominant source of error is then the mixing of the outcome n_2 with $n_2 - 1$.

(c) The finite lifetimes $T_{1,q}$ and $T_{1,m}$ are our main sources of errors as the counting probabilities $\mathcal{P}_{|n\rangle}(n)$

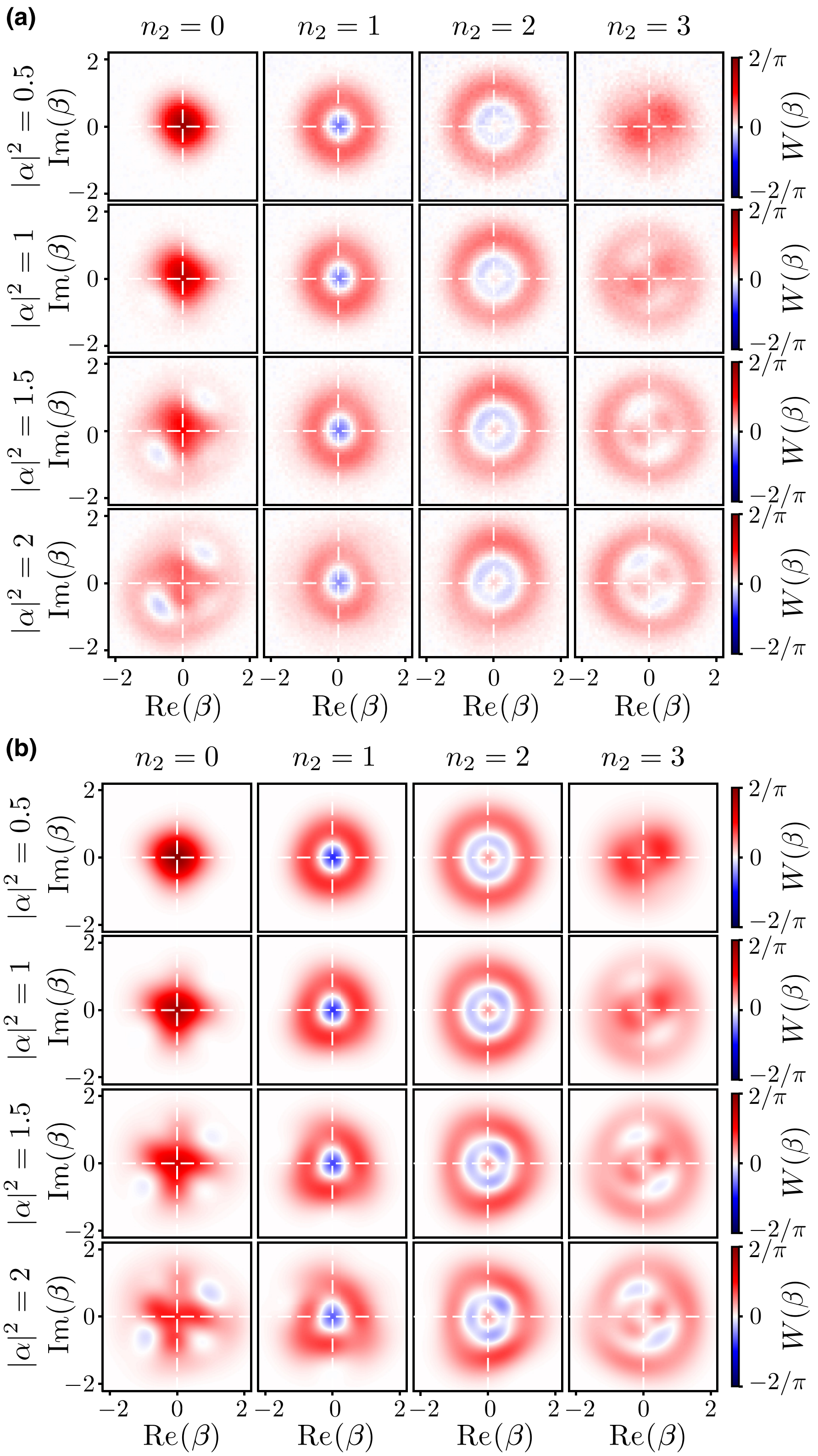

FIG. 11. Measured (a) and computed (b) Wigner functions after catching a coherent state with a mean photon number $|\alpha|^2 = 0.5$, 1, 1.5, and 2 from top to bottom, respectively, and heralding on a detected number $n_2 = 0$, 1, 2, or 3 from left to right, respectively. For each panel, the fidelity between the measured Wigner function and the predicted one does not get below 95%.

[Fig. 12(c)] and state fidelities $\mathcal{F}$ [Fig. 13(c)] get close to 1 when both $T_{1,q}$ and $T_{1,m}$ increase. If both $T_{1,q}$ and $T_{1,m}$ increase by an order of magnitude, the success probability will not get below 85% for all outcomes [Fig. 12(c)]. The QNDness is more demanding and one would need to increase by more than 2 orders of magnitude the lifetimes in order to get fidelities beyond 80% [insets of Figs. 13(a)–13(c)]. Note that current state of the art in

TABLE II. Fidelities $\mathcal{F}$ between the measured collapsed quantum states ρ and the ideal quantum states $\rho_{n_2} = |\psi_{n_2}\rangle\langle\psi_{n_2}|$ for various outcomes n_2 and various mean photon numbers $|\alpha|^2$.

$\mathcal{F}(\rho, \rho_{n_2})$	$n_2 = 0$	$n_2 = 1$	$n_2 = 2$	$n_2 = 3$
$\lvert\alpha\rvert^2 = 0.5$	86%	52%	32%	4.9%
$\lvert\alpha\rvert^2 = 1$	77%	50%	34%	11%
$\lvert\alpha\rvert^2 = 1.5$	58%	48%	38%	18%
$\lvert\alpha\rvert^2 = 2$	39%	42%	37%	22%

three-dimensional (3D) cavities and new materials demonstrates lifetimes indeed larger than 2 orders of magnitude [64,65].

(d) Our device does not seem to be limited by thermal excitations [Figs. 12(d) and 13(d)].

(e) A more faithful qubit readout would not bring significant improvements in the success probabilities and QNDness [Figs. 12(f) and 13(f)].

(f) The memory self-Kerr rate K does not seem to affect the success probabilities and QNDness (not shown). Indeed, the Fock states are eigenstates of the self-Kerr term. However, the additional self-Kerr rate K_e when the qubit is in $|e\rangle$ has an important impact [Figs. 12(e) and 13(e)]. During the interaction time T_k of question Q_k, the qubit acquires an additional parasitic phase $n^2 K_e T_k$ for each Fock state $|n\rangle$. Therefore, for $n \geq 1$ and each question Q_k, the qubit phase does not end up in the right value, which undermines the photon number encoding. As long as $n^2 K_e 2\pi/(\chi 2^k) \ll 1$, this effect can be neglected. For our device, it translates into $n \ll 3.7$. This square dependence on the photon number n is the main limitation of this scheme for increasing the maximal number of photons the detector can resolve. Similar to Ref. [66], we compute the rate K_e using perturbation theory to the fourth order in the transverse coupling strength

$$g = \sqrt{\chi \Delta(\Delta - K_q)/(2K_q)}.$$

It is obtained as a function of the detuning $\Delta = \omega_m - \omega_q$, transmon anharmonicity $-K_q = -E_C/\hbar$ and dispersive shift χ

$$K_e = \frac{\chi^2}{K_q} \frac{[2\Delta^3 - (\Delta - K_q)^3]}{2\Delta(\Delta - K_q)(\Delta + K_q)}. \tag{H1}$$

It is then possible to reduce K_e considerably while preserving the behavior of the device for large photon numbers by careful optimization of the device

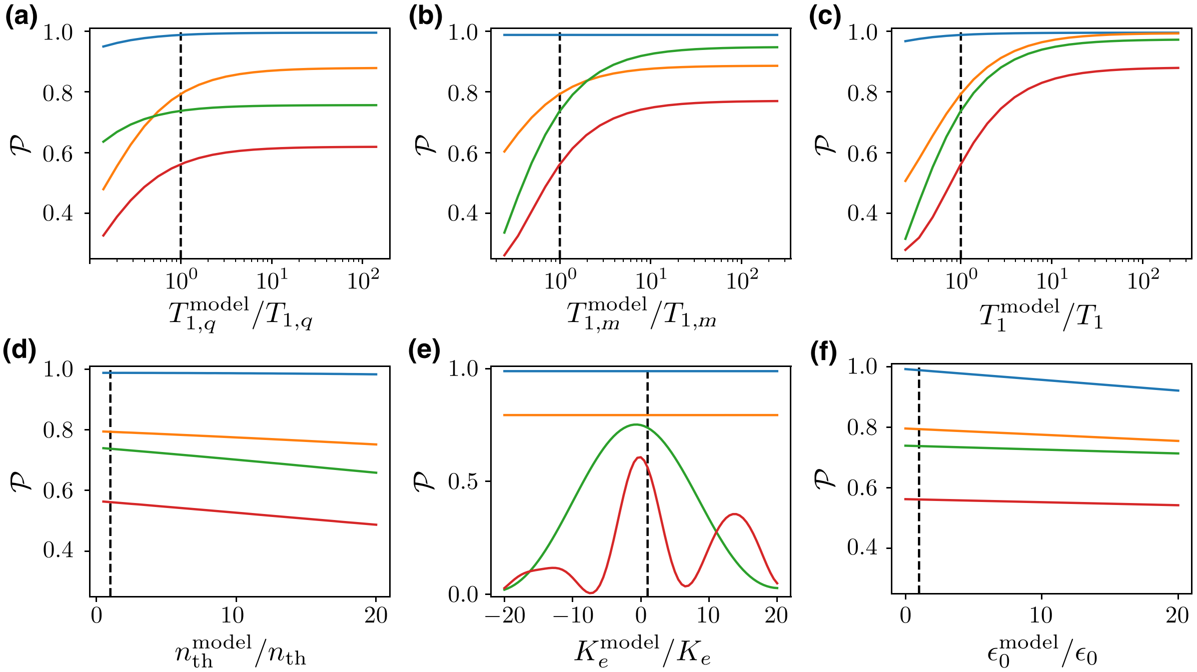

FIG. 12. Success probabilities $\mathcal{P}_{|0\rangle}(0)$ (blue), $\mathcal{P}_{|1\rangle}(1)$ (orange), $\mathcal{P}_{|2\rangle}(2)$ (green), and $\mathcal{P}_{|3\rangle}(3)$ (red) as a function of the ratio between the parameter in the model and the same parameter in experiment. All curves are calculated in the case of an initial coherent state of amplitude $|\alpha| = \sqrt{0.5}$. Vertical dashed lines indicate the result of the model for the actual experiment. Each panel probes the errors coming from (a) the qubit relaxation time $T_{1,q}$, (b) the memory relaxation time $T_{1,m}$, (c) both qubit and memory relaxation times $T_1 = (T_{1,q}, T_{1,m})$, (d) qubit and memory thermal population, (e) additional Kerr rate K_e when the qubit is excited, and (f) readout error ϵ_0.

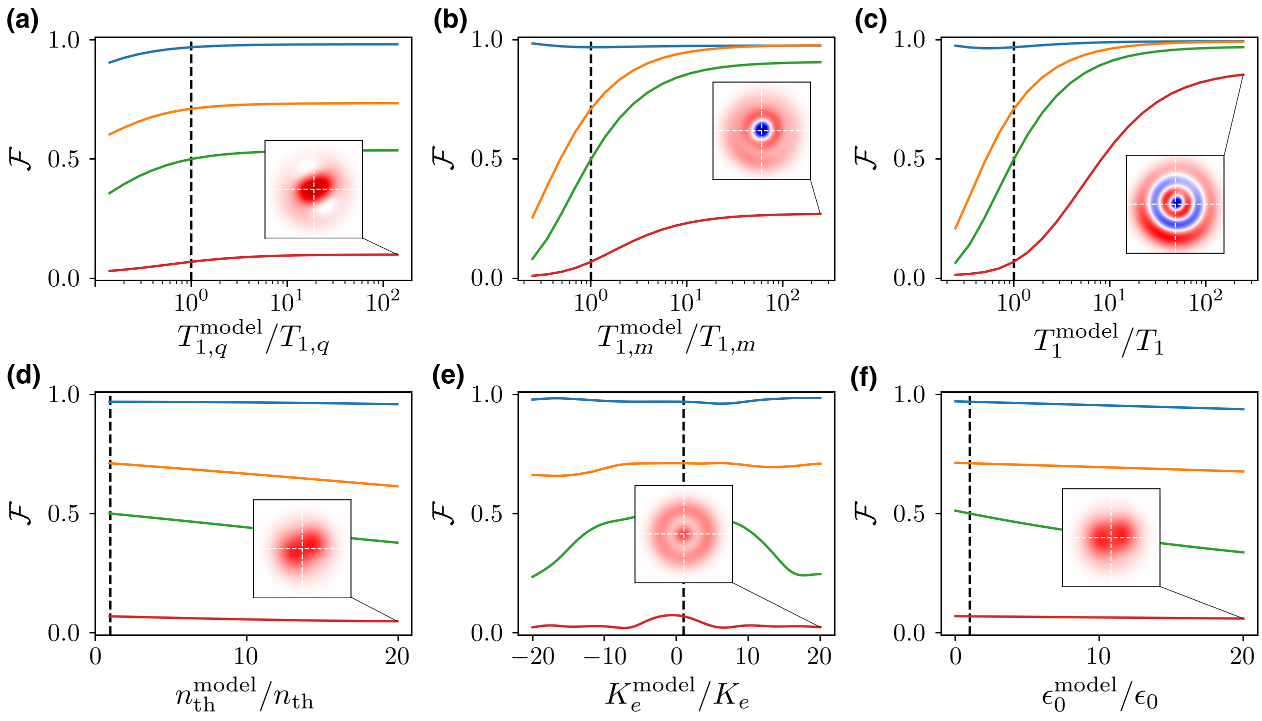

FIG. 13. QNDness of the detector. Fidelity $\mathcal{F}$ between the quantum state ρ predicted by our model and the ideal projected state ρ_{n_2} after catching a coherent state of amplitude $|\alpha| = \sqrt{0.5}$ for the outcomes $n_2 = 0$ (blue), $n_2 = 1$ (orange), $n_2 = 2$ (green), and $n_2 = 3$ (red). Each panel addresses the same parameter as in Fig. 12. Insets are the Wigner functions heralded on the counter outcome $n_2 = 3$ for the maximal value of the model parameter. Note that on the top panels, the maximal value improves QNDness while it deteriorates it for bottom panels.

parameters. For example, setting the detuning accurately to $\Delta = K_q/(1 - \sqrt[3]{2})$ cancels the rate K_e completely.

[1] R. H. Hadfield, Single-photon detectors for optical quantum information applications, Nat. Photonics **3**, 696 (2009).

[2] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deléglise, U. Busk Hoff, M. Brune, J. M. Raimond, and S. Haroche, Quantum jumps of light recording the birth and death of a photon in a cavity, Nature **446**, 297 (2007).

[3] C. Guerlin, J. Bernu, S. Deléglise, C. Sayrin, S. Gleyzes, S. Kuhr, M. Brune, J. M. Raimond, and S. Haroche, Progressive field-state collapse and quantum non-demolition photon counting, Nature **448**, 889 (2007).

[4] B. R. Johnson, M. D. Reed, A. A. Houck, D. I. Schuster, L. S. Bishop, E. Ginossar, J. M. Gambetta, L. Dicarlo, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Quantum non-demolition detection of single microwave photons in a circuit, Nat. Phys. **6**, 663 (2010).

[5] P. J. Leek, M. Baur, J. M. Fink, R. Bianchetti, L. Steffen, S. Filipp, and A. Wallraff, Cavity Quantum Electrodynamics with Separate Photon Storage and Qubit Readout Modes, Phys. Rev. Lett. **104**, 100504 (2010).

[6] L. Sun, A. Petrenko, Z. Leghtas, B. Vlastakis, G. Kirchmair, K. M. Sliwa, A. Narla, M. Hatridge, S. Shankar, J. Blumoff, L. Frunzio, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Tracking photon jumps with repeated quantum non-demolition parity measurements, Nature **511**, 444 (2014).

[7] G. Romero, J. J. García-Ripoll, and E. Solano, Microwave Photon Detector in Circuit QED, Phys. Rev. Lett. **102**, 173602 (2009).

[8] F. Helmer, M. Mariantoni, E. Solano, and F. Marquardt, Quantum nondemolition photon detection in circuit QED and the quantum Zeno effect, Phys. Rev. A **79**, 052115 (2009).

[9] K. Koshino, K. Inomata, T. Yamamoto, and Y. Nakamura, Implementation of an Impedance-Matched Λ System by Dressed-State Engineering, Phys. Rev. Lett. **111**, 153601 (2013).

[10] S. R. Sathyamoorthy, L. Tornberg, A. F. Kockum, B. Q. Baragiola, J. Combes, C. M. Wilson, T. M. Stace, and G. Johansson, Quantum Nondemolition Detection of a Propagating Microwave Photon, Phys. Rev. Lett. **112**, 093601 (2014).

[11] B. Fan, G. Johansson, J. Combes, G. J. Milburn, and T. M. Stace, Nonabsorbing high-efficiency counter for itinerant microwave photons, Phys. Rev. B **90**, 035132 (2014).

[12] O. Kyriienko and A. S. Sørensen, Continuous-Wave Single-Photon Transistor Based on a Superconducting Circuit, Phys. Rev. Lett. **117**, 140503 (2016).

[13] S. R. Sathyamoorthy, T. M. Stace, and G. Johansson, Detecting itinerant single microwave photons, C. R. Phys. **17**, 756 (2016).

[14] X. Gu, A. Frisk, A. Miranowicz, Y.-X. Liu, and F. Nori, Microwave photonics with superconducting quantum circuits, Phys. Rep. **718–719**, 1 (2017).

[15] C. H. Wong and M. G. Vavilov, Quantum efficiency of a single microwave photon detector based on a semiconductor double quantum dot, Phys. Rev. A **95**, 012325 (2017).

[16] J. Leppäkangas, M. Marthaler, D. Hazra, S. Jebari, R. Albert, F. Blanchet, G. Johansson, and M. Hofheinz, Multiplying and detecting propagating microwave photons using inelastic Cooper-pair tunneling, Phys. Rev. A **97**, 013855 (2018).

[17] B. Royer, A. L. Grimsmo, A. Choquette-Poitevin, and A. Blais, Itinerant Microwave Photon Detector, Phys. Rev. Lett. **120**, 203602 (2018).

[18] Y. F. Chen, D. Hover, S. Sendelbach, L. Maurer, S. T. Merkel, E. J. Pritchett, F. K. Wilhelm, and R. McDermott, Microwave Photon Counter Based on Josephson Junctions, Phys. Rev. Lett. **107**, 217401 (2011).

[19] K. Inomata, Z. Lin, K. Koshino, W. D. Oliver, J. S. Tsai, T. Yamamoto, and Y. Nakamura, Single microwave-photon detector using an artificial Λ-type three-level system, Nat. Commun. **7**, 12303 (2016).

[20] J. C. Besse, S. Gasparinetti, M. C. Collodo, T. Walter, P. Kurpiers, M. Pechal, C. Eichler, and A. Wallraff, Single-Shot Quantum Nondemolition Detection of Individual Itinerant Microwave Photons, Phys. Rev. X **8**, 21003 (2018).

[21] S. Kono, K. Koshino, Y. Tabuchi, A. Noguchi, and Y. Nakamura, Quantum non-demolition detection of an itinerant microwave photon, Nat. Phys. **14**, 546 (2018).

[22] A. Narla, S. Shankar, M. Hatridge, Z. Leghtas, K. M. Sliwa, E. Zalys-Geller, S. O. Mundhada, W. Pfaff, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret, Robust Concurrent Remote Entanglement between two Superconducting Qubits, Phys. Rev. X **6**, 031036 (2016).

[23] R. Lescanne, S. Deléglise, E. Albertinale, U. Réglade, T. Capelle, E. Ivanov, T. Jacqmin, Z. Leghtas, and E. Flurin, Irreversible Qubit-Photon Coupling for the Detection of Itinerant Microwave Photons, Phys. Rev. X **10**, 021038 (2019).

[24] A. M. Sokolov and F. K. Wilhelm, A superconducting detector that counts microwave photons up to two, arXiv:2003.04625 [quant-ph] (2020).

[25] A. L. Grimsmo, B. Royer, J. M. Kreikebaum, Y. Ye, K. O'Brien, I. Siddiqi, and A. Blais, Quantum metamaterial for nondestructive microwave photon counting, arXiv:2005.06483 [quant-ph] (2020).

[26] N. Bergeal, F. Schackert, M. Metcalfe, R. Vijay, V. E. Manucharyan, L. Frunzio, D. E. Prober, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret, Phase-preserving amplification near the quantum limit with a Josephson ring modulator, Nature **465**, 64 (2010).

[27] N. Roch, E. Flurin, F. Nguyen, P. Morfin, P. Campagne-Ibarcq, M. H. Devoret, and B. Huard, Widely Tunable, Nondegenerate Three-Wave Mixing Microwave Device Operating near the Quantum Limit, Phys. Rev. Lett. **108**, 147701 (2012).

[28] T. Peronnin, D. Marković, Q. Ficheux, and B. Huard, Sequential Dispersive Measurement of a Superconducting Qubit, Phys. Rev. Lett. **124**, 180502 (2020).

[29] Y. Yin, Y. Chen, D. Sank, P. J. J. O'Malley, T. C. White, R. Barends, J. Kelly, E. Lucero, M. Mariantoni, A. Megrant, C. Neill, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Catch and Release of Microwave Photon States, Phys. Rev. Lett. **110**, 107001 (2013).

[30] J. Wenner, Y. Yin, Y. Chen, R. Barends, B. Chiaro, E. Jeffrey, J. Kelly, A. Megrant, J. Y. Mutus, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, T. C. White, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Catching Time-Reversed Microwave Coherent State Photons with 99.4% Absorption Efficiency, Phys. Rev. Lett. **112**, 210501 (2014).

[31] E. Flurin, Ph.D. thesis, School École Normale Supérieure, 2014.

[32] C. J. Axline, L. D. Burkhart, W. Pfaff, M. Zhang, K. Chou, P. Campagne-Ibarcq, P. Reinhold, L. Frunzio, S. M. Girvin, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, On-demand quantum state transfer and entanglement between remote microwave cavity memories, Nat. Phys. **14**, 705 (2018).

[33] Y. P. Zhong, H. S. Chang, K. J. Satzinger, M. H. Chou, A. Bienfait, C. R. Conner, Dumur,J. Grebel, G. A. Peairs, R. G. Povey, D. I. Schuster, and A. N. Cleland, Violating Bell's inequality with remotely connected superconducting qubits, Nat. Phys. **15**, 741 (2019).

[34] P. Campagne-Ibarcq, E. Zalys-Geller, A. Narla, S. Shankar, P. Reinhold, L. Burkhart, C. Axline, W. Pfaff, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret, Deterministic Remote Entanglement of Superconducting Circuits through Microwave Two-Photon Transitions, Phys. Rev. Lett. **120**, 200501 (2018).

[35] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J. C. Besse, S. Gasparinetti, A. Blais, and A. Wallraff, Deterministic quantum state transfer and remote entanglement using microwave photons, Nature **558**, 264 (2018).

[36] A. N. Korotkov, Flying microwave qubits with nearly perfect transfer efficiency, Phys. Rev. B **84**, 014510 (2011).

[37] E. Flurin, N. Roch, J. D. Pillet, F. Mallet, and B. Huard, Superconducting Quantum Node for Entanglement and Storage of Microwave Radiation, Phys. Rev. Lett. **114**, 1 (2015).

[38] D. I. Schuster, A. A. Houck, J. A. Schreier, A. Wallraff, J. M. Gambetta, A. Blais, L. Frunzio, J. Majer, B. Johnson, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Resolving photon number states in a superconducting circuit, Nature **445**, 515 (2007).

[39] D. T. McClure, H. Paik, L. S. Bishop, M. Steffen, J. M. Chow, and J. M. Gambetta, Rapid Driven Reset of a Qubit Readout Resonator, Phys. Rev. Appl. **5**, 11001 (2016).

[40] S. Haroche, M. Brune, and J. Raimond, Measuring photon numbers in a cavity by atomic interferometry: Optimizing the convergence procedure, J. Phys. II **2**, 659 (1992).

[41] R. Heeres, P. Reinhold, and R. Schoelkopf, (private communication).

[42] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits, Phys. Rev. Lett. **103**, 110501 (2009).

[43] C. S. Wang, J. C. Curtis, B. J. Lester, Y. Zhang, Y. Y. Gao, J. Freeze, V. S. Batista, P. H. Vaccaro, I. L. Chuang, L. Frunzio, L. Jiang, S. M. Girvin, and R. J. Schoelkopf, Efficient Multiphoton Sampling of Molecular Vibronic Spectra on a Superconducting Bosonic Processor, Phys. Rev. X **10**, 021060 (2020).

[44] M. Khezri, E. Mlinar, J. Dressel, and A. N. Korotkov, Measuring a transmon qubit in circuit QED: Dressed squeezed states, Phys. Rev. A **94**, 12347 (2016).

[45] L. G. Lutterbach and L. Davidovich, Method for Direct Measurement of the Wigner Function in Cavity QED and Ion Traps, Phys. Rev. Lett. **78**, 2547 (1997).

[46] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, Direct Measurement of the Wigner Function of a One-Photon Fock State in a Cavity, Phys. Rev. Lett. **89**, 200402 (2002).

[47] B. Vlastakis, G. Kirchmair, Z. Leghtas, S. E. Nigg, L. Frunzio, S. M. Girvin, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Deterministically encoding quantum information using 100-photon Schrödinger cat states, Science **342**, 607 (2013).

[48] P. E. M. F. Mendonça, R. d. J. Napolitano, M. A. Marchiolli, C. J. Foster, and Y.-C. Liang, Alternative fidelity measure between quantum states, Phys. Rev. A **78**, 052330 (2008).

[49] J. A. Miszczak, Z. Puchała, P. Horodecki, A. Uhlmann, and K. Zyczkowski, Sub- and super-fidelity as bounds for quantum fidelity, Quantum Info. Comput. **9**, 103 (2009).

[50] J.-C. Besse, S. Gasparinetti, M. C. Collodo, T. Walter, A. Remm, J. Krause, C. Eichler, and A. Wallraff, Parity Detection of Propagating Microwave Fields, Phys. Rev. X **10**, 11046 (2020).

[51] S. J. Dolinar, An optimum receiver for the binary coherent state quantum channel, MIT Res. Lab. Electron. Q. Prog. Rep. **111**, 115 (1973).

[52] https://github.com/Quantum-Circuit-Group/photocounting-OPX.

[53] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, A near–quantum-limited Josephson traveling-wave parametric amplifier, Science **350**, 307 (2015).

[54] D. Sank *et al.*, Measurement-Induced State Transitions in a Superconducting Qubit: Beyond the Rotating Wave Approximation, Phys. Rev. Lett. **117**, 190503 (2016).

[55] R. Lescanne, L. Verney, Q. Ficheux, M. H. Devoret, B. Huard, M. Mirrahimi, and Z. Leghtas, Escape of a Driven Quantum Josephson Circuit Into Unconfined States, Phys. Rev. Appl. **11**, 14030 (2019).

[56] S. Touzard, A. Kou, N. E. Frattini, V. V. Sivak, S. Puri, A. Grimm, L. Frunzio, S. Shankar, and M. H. Devoret, Gated Conditional Displacement Readout of Superconducting Qubits, Phys. Rev. Lett. **122**, 80502 (2019).

[57] J. Ikonen, J. Goetz, J. Ilves, A. Keränen, A. M. Gunyho, M. Partanen, K. Y. Tan, D. Hazra, L. Grönberg, V. Vesterinen, S. Simbierowicz, J. Hassel, and M. Möttönen, Qubit Measurement by Multichannel Driving, Phys. Rev. Lett. **122**, 80503 (2019).

[58] R. Dassonneville, T. Ramos, V. Milchakov, L. Planat, É. Dumur, F. Foroughi, J. Puertas, S. Leger, K. Bharadwaj, J. Delaforce, C. Naud, W. Hasch-Guichard, J. J. García-Ripoll, N. Roch, and O. Buisson, Fast High-Fidelity Quantum Nondemolition Qubit Readout via a Nonperturbative Cross-Kerr Coupling, Phys. Rev. X **10**, 11045 (2020).

[59] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, Phys. Rev. A **91**, 22118 (2015).

[60] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits, Phys. Rev. Appl. **7**, 054020 (2017).

[61] M. Fliess, J. Lévine, P. Martin, and P. Rouchon, Flatness and defect of non-linear systems: Introductory theory and examples, Int. J. Control **61**, 1327 (1995).

[62] P. Campagne-Ibarcq, Ph.D. thesis, School École Normale Supérieure (ENS), 2015.

[63] S. Krämer, D. Plankensteiner, L. Ostermann, and H. Ritsch, QuantumOptics.jl: A Julia framework for simulating open quantum systems, Comput. Phys. Commun. **227**, 109 (2018).

[64] M. Reagor, W. Pfaff, C. Axline, R. W. Heeres, N. Ofek, K. Sliwa, E. Holland, C. Wang, J. Blumoff, K. Chou, M. J. Hatridge, L. Frunzio, M. H. Devoret, L. Jiang, and R. J. Schoelkopf, Quantum memory with millisecond coherence in circuit QED, Phys. Rev. B **94**, 014506 (2016).

[65] A. P. M. Place, L. V. H. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, S. Sussman, G. Cheng, T. Madhavan, H. K. Babla, B. Jaeck, A. Gyenis, N. Yao, R. J. Cava, N. P. de Leon, and A. A. Houck, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, arXiv:2003.00024 [quant-ph] (2020).

[66] M. Elliott, J. Joo, and E. Ginossar, Designing Kerr interactions using multiple superconducting qubit types in a single circuit, New J. Phys. **20**, 023037 (2018).