# Raman quantum memory with built-in suppression of four-wave-mixing noise

S. E. Thomas,<sup>1,2</sup> T. M. Hird,<sup>1,3</sup> J. H. D. Munns,<sup>1,2</sup> B. Brecht,<sup>1,4</sup> D. J. Saunders,<sup>1</sup>

J. Nunn,<sup>5</sup> I. A. Walmsley,<sup>1,2</sup> and P. M. Ledingham<sup>1</sup>

<sup>1</sup>Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

<sup>2</sup>QOLS, Department of Physics, Imperial College London, London SW7 2BW, United Kingdom

<sup>3</sup>Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom <sup>4</sup>Applied Physics, University of Paderborn, Warburgerstrasse 100, 33098 Paderborn, Germany

<sup>5</sup>Centre for Photonics and Photonic Materials, University of Bath, Bath BA2 7AY, United Kingdom

(Received 26 April 2019; published 3 September 2019)

Quantum memories are essential for large-scale quantum information networks. Along with high efficiency, storage lifetime, and optical bandwidth, it is critical that the memory adds negligible noise to the recalled signal. A common source of noise in optical quantum memories is spontaneous four-wave mixing. We develop and implement a technically simple scheme to suppress this noise mechanism by means of quantum interference. Using this scheme with a Raman memory in warm atomic vapor, we demonstrate over an order of magnitude improvement in noise performance. Furthermore we demonstrate a method to quantify the remaining noise contributions and present a route to enable further noise suppression. Our scheme opens the way to quantum demonstrations using a broadband memory, significantly advancing the search for scalable quantum photonic networks.

DOI: 10.1103/PhysRevA.100.033801

## I. INTRODUCTION

An optical quantum memory is a device that can faithfully store and release quantum states of light on demand. This is a key element for future photonic quantum information protocols as a means to synchronize probabilistic processes via multiplexing, enabling secure long-distance communication through the distribution of entangled states [1]. Many impressive implementations of on-demand quantum memory (QM) protocols have been demonstrated across various platforms including warm [2-8] and cold atomic vapors [9-14]and solid-state systems [15–19]. While considerable progress has been made to reach high efficiencies [11,13,16] and long storage times [9,17], designing a memory protocol that does not add noise to the recalled signal remains a significant challenge. Common noise processes include atomic resonant fluorescence, spontaneous Raman scattering (SRS) from unpumped thermal population from the storage state, and spontaneous four-wave mixing (SFWM) [20]. Eliminating these without sacrifice to the memory lifetime or storage bandwidth has proven difficult. These spurious noise processes pollute the desired memory output field, severely limiting the signal-to-noise ratio (SNR), which, in turn, upper bounds the achievable fidelity for storing and recalling qubits [21] as well as significantly modifying the photon number statistics, i.e., the second-order autocorrelation function  $g^{(2)}$  [18,22]. Fluorescence can be reduced by operating off-resonance, and SRS from the storage state can be eliminated by near perfect optical pumping. However, removing SFWM intrinsic to broadband quantum memories remains the final hurdle.

Here we propose and demonstrate a technically simple method to suppress four-wave-mixing noise in atomic memories. Our method is widely applicable to any atomic species since it harnesses the strong linear absorption of the atomic ensemble itself. Romanov et al. have shown that SFWM noise can be suppressed via Raman absorption of the noise photon into a second atomic isotope [23,24]. Inspired by their protocol, we present a scheme for noise suppression by operating the memory at a specific detuning from resonance such that the unwanted noise field is resonant with the populated atomic transition. In this arrangement, the competing processes of noise generation by SFWM, and its absorption and dispersion by the atomic resonance, significantly suppress anti-Stokes scattering and thus the contamination of the signal field. By operating the atomic memory at a specific detuning, we utilize this built-in noise suppression (BNS) mechanism, achieving significant reduction of noise without any detrimental effects on the memory efficiency or lifetime, and without any change to the memory initialization. This is in contrast to other, more complex strategies to suppress SFWM noise such as cavity engineering [25], polarization selection rules [26], noncolinear geometry [27], and phase mismatching in dispersive media [18]. We also present a method to analyze different noise processes in quantum memories and quantify how these affect the photon number statistics of the retrieved state, thus enabling us to give a recipe for the remaining steps necessary for high-fidelity retrieval in broadband quantum memories. Therefore, our scheme holds great promise as a route towards a technically simple, noise-free quantum memory.

## **II. BNS RAMAN PROTOCOL**

The built-in noise suppression scheme is investigated in an off-resonant Raman memory in warm caesium vapor. The Raman memory protocol is based on an ensemble of atoms, each with a  $\Lambda$ -energy level configuration, which are initialized

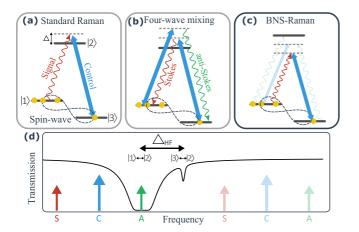


FIG. 1. (a) The Raman memory protocol uses a strong control field to drive a Raman transition and store an input signal field as a coherence across an atomic ensemble, or spin wave. (b) SFWM noise arises when the control field couples to the populated state  $|1\rangle$  and drives anti-Stokes scattering. (c) The absorptive Raman memory operates at  $\Delta = -2\Delta_{hf}$  such that the anti-Stokes field is on resonance with the  $|1\rangle \rightarrow |2\rangle$  transition, and anti-Stokes scattering is strongly suppressed. (d) Simulated absorption spectrum of warm caesium vapor in a nitrogen buffer gas at 83 °C, with 99.9% of the population in state  $|1\rangle$ . The arrows show the frequency of the signal or Stokes (S), control (C), and anti-Stokes (A) fields for the BNS (left) and STD (right) Raman protocols.

in the long-lived ground state,  $|1\rangle$  [see Fig. 1(a)]. We apply a strong control pulse in two-photon resonance with a weak signal to drive a stimulated two-photon Raman transition from  $|1\rangle$  to  $|3\rangle$ , while the two fields themselves are detuned from the excited state  $|2\rangle$  by  $\Delta$ . This coherently stores the signal field as a collective excitation, or spin wave, across the entire ensemble of atoms. To retrieve the signal, a second control pulse is applied which drives the reverse process and coherently converts the atomic spin wave back into an optical field. However, this protocol inherently suffers from SFWM noise, which inhibits operation at the quantum level [22]. The origin of SFWM noise is the unwanted coupling of the strong control field to the populated ground state,  $|1\rangle$ , which drives spontaneous anti-Stokes scattering [see Fig. 1(b)]. This creates a noisy spin wave that has significant overlap with the memory spin wave, and which is efficiently read out with the same temporal and spectral mode as the signal field. This is manifested in two ways: as a higher average retrieved signal and as an increase in the second-order intensity correlation of the signal.

By operating the Raman memory with the control field at a detuning of  $\Delta_s = -2\Delta_{\rm hf}$ , we ensure that the anti-Stokes field is on resonance with the  $|1\rangle \rightarrow |2\rangle$  transition and undergoes strong linear absorption. This is a coherent process and interferes with the SFWM generation process since both fields are at the same frequency. The destructive interference of these two pathways suppresses the generation of noise photons at the signal frequency. One may interpret this mechanism physically as the continual absorption of anti-Stokes photons that are generated by SFWM. The characteristic absorption length for the anti-Stokes field in this configuration is L/d,

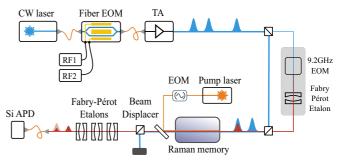


FIG. 2. Experimental setup of the Raman memory. The signal and control pulses are carved from a CW laser using a fiber integrated electro-optic interferometer (Fiber EOM) driven by rf signals from two arbitrary wave-form generators. The pulses are amplified using a tapered amplifier (TA) and split into two arms. The signal pulses pass through an EOM driven at 9.2 GHz to generate a sideband at the signal frequency, and delayed in time to overlap with the strong control pulses. The pulses interact with an ensemble of warm caesium atoms, which is heated to  $83 \,^{\circ}$ C and initialized in  $|1\rangle$  by a counterpropagating optical pumping laser. After the memory, the control pulse is filtered from the signal using a beam displacer and Fabry-Pérot etalons, before the signal is detected using a silicon single-photon avalanche photodiode (Si APD).

where L is the length of the atomic medium and d is the on-resonance optical depth of the atomic ensemble. This length scale is typically less than 100  $\mu$ m, and therefore any anti-Stokes photon will be absorbed within a short distance, meaning that the spin wave it generates will be localized. Thus SFWM cannot lead to a delocalized excitation over the entire atomic ensemble of length L, in contrast to the read-in and read-out memory interactions. The spatial confinement of the noisy spin wave means that the overlap with the memory spin-wave mode is greatly diminished. This process is akin to the dissipative quantum Zeno effect [28], in that the absorption of a noise photon acts as a measurement process which prevents noise excitations from collectively building up over the memory interaction length.

## **III. RESULTS**

A schematic of the experimental setup is shown in Fig. 2 and details are given in the Appendix. We operate the memory in two configurations: (1)  $\Delta_s = -2\Delta_{\rm hf}$ , where the anti-Stokes field is on resonance with the atomic transition and therefore strongly absorbed (BNS-Raman), (2)  $\Delta_s = +2\Delta_{\rm hf}$ , where the coupling strength of the Raman memory interaction is the same but there is no atomic suppression of the SFWM noise, i.e., the "standard" Raman memory (STD-Raman).

We define the memory efficiency as the ratio of the integrated intensity of the retrieved signal to the input signal, and this quantity contains contributions from the desired memory interaction as well as the unwanted four-wave-mixing (FWM) gain. Figure 3 shows the memory efficiency as a function of control pulse energy in these two regimes, and a typical arrival time histogram of the memory interaction measured on the single-photon avalanche photodiodes (APDs). The measured memory efficiency is lower in the BNS-Raman case compared to the STD-Raman memory and there are

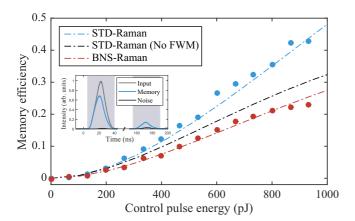


FIG. 3. Memory efficiency as a function of control pulse energy. The error bars, estimated from Poissonian errors on the number of detection events, are within the marker size. The dashed lines are the theoretical predictions from our numerical simulations for the STD-Raman memory with (blue, upper) and without (black, middle) four-wave mixing, and the BNS-Raman memory (red, lower). The inset shows typical arrival time histograms for the input signal (control pulses off), the memory interaction (signal and control pulses), and noise (control only, no signal). The shaded gray regions show the integration windows which are set to 35 ns. The memory retrieval time is 150 ns, and the input photon number for the STD-(BNS-)Raman memory is 3.5 (3.2).

two main reasons for this. First, we have suppressed fourwave mixing and hence have removed the gain process which amplifies the input signal. The black line in Fig. 3 shows the predicted memory efficiency in the STD-Raman memory if we artificially turn off four-wave mixing, and we see that is it significantly lower than the measured efficiency. Second, the control field is much closer in frequency to the populated  $|1\rangle \rightarrow |2\rangle$  transition for BNS-Raman [9.2 GHz compared to 27.6 GHz for STD-Raman; see Fig. 1(d)], which leads to more linear absorption of the control field in the atomic medium. For the same control pulse energy before the memory, the efficiency is lower for BNS-Raman than STD-Raman due to partial absorption of the control pulse. However, this is not a fundamental limitation and, with sufficient pulse energy and appropriate shaping of the control pulse, the efficiency can, in principle, reach 100% in both configurations [29].

The memory lifetime is  $(294 \pm 26)$  ns for the STD-Raman memory and  $(625 \pm 36)$  ns for the BNS-Raman memory. The factor limiting these timescales is the extinction ratio that can be achieved with the intensity modulator switching the optical pumping light, where any residual leakage during storage continues to pump the atoms and causes the spin wave to be depleted. This extinction ratio varies over time, which causes the discrepancy in the storage time between the two experiments. A better modulator would allow memory lifetimes limited in this case by the diffusion of atoms out of the control beam, thus similar to the previously demonstrated 2  $\mu$ s lifetime [22].

To confirm the noise suppression, we measure the average number of noise photons generated per control pulse,  $N_{\text{noise}}$ , as shown in Fig. 4. The amount of noise is strongly suppressed in the BNS-Raman scheme. At a control pulse

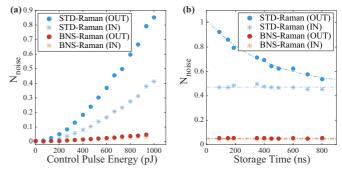


FIG. 4. The measured noise photons per pulse,  $N_{\text{noise}}$ , in the input (circles) and storage (stars) time bins as a function of (a) control pulse energy and (b) memory storage time for the STD- (blue, upper) and the BNS- (red, lower) Raman memory. The storage time for (a) is 70 ns and the control pulse energy for (b) is 930 pJ. The error bars, estimated from Poissonian errors on the number of detection events, are within the marker size.

energy of 930 pJ and storage time of 70 ns, the unconditional noise level is reduced from  $N_{\text{noise}}^{(\text{STD})} = 0.793(2)$  to  $N_{\text{noise}}^{(\text{BNS})} = 0.0467(5)$  photons per pulse—a decrease by a factor of 17. The memory efficiency for these parameters is  $\eta^{(\text{STD})} =$ 42.8(2)% and  $\eta^{(\text{BNS})} = 23.0(1)\%$ , respectively, and therefore the noise-to-efficiency ratio,  $\mu_1 = N_{\text{noise}}/\eta$ , sees a significant decrease from  $\mu_1^{(\text{STD})} = 1.85(9)$  to  $\mu_1^{(\text{BNS})} = 0.20(2)$ . This demonstrates that the BNS-Raman scheme is a powerful, technically simple method to reduce the noise in the Raman memory, without detriment to the memory efficiency or lifetime.

We gain further insight by looking at how the noise level scales with the memory readout time. For the STD-Raman memory, the noise in the retrieval time bin decreases with storage time with a decay constant of  $(380 \pm 36)$  ns. This indicates that the noise process involves an atomic coherence which decays at a similar rate as the memory efficiencyconsistent with four-wave-mixing noise. In contrast, for the BNS-Raman memory, the noise is independent of the readout time. We also see that in the STD-Raman memory, the noise is significantly higher in the retrieval time bin than in the input time bin. This is again consistent with SFWM noise which builds up with subsequent applications of the control pulse due to the generation and partial retrieval of a spin-wave excitation with each pulse [22]. Other sources of noise such as spontaneous scattering due to imperfect optical pumping and fluorescence would generate the same amount of noise photons on every application of the control pulses and would therefore be equal in all time bins. For the BNS-Raman memory, the noise is almost identical for the storage and retrieval time bins. These results give a strong indication that we have suppressed the four-wave-mixing process and, for a control pulse energy of 930 pJ, have residual noise sources remaining of  $\sim 0.05$  photons per pulse.

### **IV. NOISE CHARACTERIZATION**

To further characterize the output of the memory and confirm the suppression of the SFWM noise, we consider the second-order autocorrelation function of the output photonic state,  $g_{out}^{(2)}$ , which allows us to investigate the different noise sources in more detail. This is defined as [30]

$$g_{\text{out}}^{(2)}(\tau) = \frac{\iint dt dt' \langle S_{\text{out}}^{\dagger}(t) S_{\text{out}}^{\dagger}(t+\tau) S_{\text{out}}(t'+\tau) S_{\text{out}}(t') \rangle}{\left(\int dt \langle S_{\text{out}}^{\dagger}(t) S_{\text{out}}(t) \rangle\right)^2}$$

where  $S_{\text{out}}$  is the output signal mode operator and  $g^{(2)}(\tau =$ (0) < 1 signifies a field with nonclassical statistics. We consider the average number of noise photons arriving at the detector for three noise processes that could contribute to the photon number statistics: (1) spontaneous Raman scattering, N<sub>SRS</sub>, either from FWM or from spontaneous Stokes scattering from the unpumped thermal population in  $|3\rangle$  with  $g_{SRS}^{(2)} = 2$ ; (2) broadband collisional-induced fluorescence,  $N_{\rm F}$ , that is not sufficiently filtered from detection. We note that in general. the noise due to fluorescence will have multimode thermal statistics, with  $1 \leq g_{\rm F}^{(2)} \leq 2$  depending on how many modes are collected. We assume that the fluorescence noise that we collect has single-mode thermal statistics with  $g_{\rm F}^{(2)} = 2$  since our detection is after significant spectral filtering and within a time window of 35 ns, which is comparable to the fluorescence timescale; and (3) control field leakage,  $N_{\rm L}$ , which we consider to be zero, evidenced by measuring detection dark counts when the memory medium is removed. We derive an expression for the  $g^{(2)}(0)$  of the retrieved state as a function of the output photon number,  $N_{out} = \eta N_{in}$ , taking into account these different noise processes, and predict

$$g_{\text{out}}^{(2)} = 1 + \frac{aN_{\text{out}}^2 + 2N_{\text{SRS}}N_{\text{out}} + N_{\text{SRS}}^2 + N_{\text{F}}^2}{(N_{\text{out}} + N_{\text{SRS}} + N_{\text{F}})^2}, \qquad (1)$$

where  $a = g_{in}^{(2)} \mathcal{G}_{ss}/\eta^2 - 1$ , and  $\mathcal{G}_{ss}$  is the integrated Green's function kernel describing the linear mapping from the input signal field to the retrieved signal field. For more details, see the Appendix.

To quantify the relative contributions of these different noise sources, we measure the second-order autocorrelation of the retrieved optical field for coherent state inputs  $g_{in}^{(2)} = 1$ with average photon numbers varying from  $N_{in} \sim 0.5$  to 80, as well as measuring the autocorrelation of the noise with  $N_{in} =$ 0. The output signal is sent to a Hanbury-Brown-Twiss setup comprised of a half waveplate, a polarizing beam splitter, and two fiber-coupled APDs. The correlations between the two detectors are measured using a time tagger (Swabian Instruments Time Tagger 20) to calculate  $g_{out}^{(2)}$ .

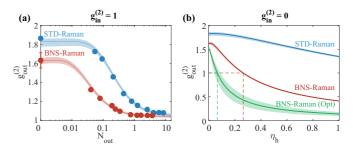


FIG. 5. (a) The measured  $g^{(2)}$  of the retrieved state from the memory as a function of the retrieved photon number,  $N_{out}$ , for the BNS- and STD-Raman memory. The control pulse energy is 330 pJ and the storage time is 150 ns. The solid lines are the fit to data using Eq. (1), and the shaded regions indicate the 95% confidence intervals on the fit. (b) The predicted  $g_{out}^{(2)}$  for a single-photon input with  $g_{in}^{(2)} = 0$  as a function of the input photon number or heralding efficiency,  $\eta_h$ , using the fitted parameters. The green line indicates the  $g^{(2)}$  that could be reached by optimizing the BNS scheme via improvement of the optical pumping.

 $3.8(5) \times 10^{-3}$ . Therefore, along with the significant decrease in the average noise, the BNS case presents a change in the SRS-to-fluorescence ratio (~3, compared to ~9 for the Raman memory), resulting in a noticeable reduction of  $g_{out}^{(2)}$  at zero input, while the increased SNR allows for faster scaling to  $g_{in}^{(2)}$  as the input photon number is increased.

To distinguish whether the remaining noise from spontaneous Raman scattering, N<sub>SRS</sub>, is due to four-wave mixing or Stokes scattering due to imperfect optical pumping, we measure the noise as a function of the optical pumping efficiency. We find that the noise decreases linearly with the amount of residual population in the storage state, which is consistent with noise from spontaneous scattering from the unpumped population. We measure that we can decrease the total noise,  $N_{\text{noise}} = N_{\text{F}} + N_{\text{SRS}}$ , to  $N_{\text{noise}} = 6.6(2) \times 10^{-3}$ with higher optical pumping power (see the Appendix). Even for perfect optical pumping, we predict  $N_{\text{noise}} = 5(2) \times 10^{-3}$ , which agrees well with the extracted value of  $N_F^{(BNS)} =$  $3.8(5) \times 10^{-3}$ . We therefore conclude that the residual noise is a combination of fluorescence noise and spontaneous Raman scattering due to imperfect optical pumping, and that we have successfully eliminated SFWM noise.

To explore the efficacy of this scheme for enabling quantum-level storage, we use Eq. (1) together with the fitting parameters from the weak coherent state data to predict the output photon statistics for the case when the input is a single-photon Fock state with  $g_{in}^{(2)} = 0$ . Figure 5(b) shows the predicted  $g_{out}^{(2)}$  as a function of the probability for a single photon to arrive at the input of the quantum memory,  $\eta_h$ . The STD-Raman memory is unable to produce a nonclassical output state, even for unit probability, due to the significant noise contribution. In contrast, the BNS-Raman case with the large reduction FWM noise is able to output nonclassical states for heralding efficiencies exceeding (26.4 ± 0.5)%. This is well within the performance parameters of existing technologies, with heralding efficiencies as high as 87% possible [31].

By measuring the decrease in noise as we increased the optical pumping power, we have determined that we could reduce the total noise to  $N_{\text{noise}} = 6.6(2) \times 10^{-3}$ , of which approximately half is due to fluorescence  $[N_{\text{(BNS)}}^{(\text{BNS)}} = 3.8(5) \times$  $10^{-3}$ ] and half is due to imperfect optical pumping. Furthermore, improved pump switching extinction will prevent spin-wave depletion during storage, improving our efficiency from 10.2% to 12.7% for this control pulse energy, with the same level of noise. These changes would give  $\mu_1^{opt} =$ 0.052(3), or an upper bound on the conditional fidelity of a retrieved qubit of around  $\mathcal{F} = 0.95$  [21]. This would also yield the green line in Fig. 5(b), i.e., a drop in the requisite single-photon heralding efficiency for nonclassical readout to  $(6.5 \pm 0.6)$ %. In this case, a photon from a deterministic single-photon source could be retrieved from the memory with  $g_{out}^{(2)} = 0.14 \pm 0.02$ , significantly below the nonclassical threshold. Further improvement of the optical pumping, and better spectral filtering to remove fluorescence noise, would allow for even more dramatic improvements.

#### V. CONCLUSION

We have demonstrated a scheme for the suppression of noise in a quantum memory in a warm atomic vapor by means of coherent destructive interference of SFWM and absorption. We have shown that this built-in noise suppression offers reduction in SFWM noise in a Raman memory by more than an order of magnitude, reaching a level where nonclassical operation is possible. By quantifying different noise contributions in a quantum memory in terms of underlying physical processes, we conclude that the remaining noise is a combination of collisional-induced fluorescence and spontaneous Raman scattering from imperfect optical pumping. Further, we estimate that this technical noise could be suppressed by improved frequency filtering and by increasing the optical pumping efficiency. Our scheme is broadly applicable to any off-resonant memory protocol or any system that suffers from SFWM noise, and in particular this method would be efficacious in a cold atomic system where collisional-induced fluorescence noise is negligible. This is a technically simple and effective noise suppression scheme, which paves the way towards quantum-level storage in a long-lived, broadband, room-temperature quantum memory.

### ACKNOWLEDGMENTS

We thank K. Heshami, O. Lazo-Arjona, and J. Becker for insightful discussions. This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through the Standard Grant No. EP/J000051/1, Programme Grant No. EP/K034480/1, the EPSRC Hub for Networked Quantum Information Technologies (NQIT), and an ERC Advanced Grant (Grant No. MOQUACINO). This work has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 665148 (QCUMbER). S.E.T. and J.H.D.M. are supported by EPSRC via the Controlled Quantum Dynamics CDT under Grants No. EP/G037043/1 and No. EP/L016524/1, and T.M.H. is supported via the EPSRC Training and Skills Hub InQuBATE Grant No. EP/P510270/1. J.N. acknowledges financial support from a Royal Society University Research Fellowship (SPEARS), D.J.S. acknowledges financial support from an EU Marie Curie Fellowship No. PIIF-GA-2013-629229, and P.M.L. acknowledges financial support from an EU Marie Curie individual fellowship (Quantum BOSS) funded under H2020-EU.1.3.2. Grant Agreement ID 705278.

## APPENDIX

#### 1. Equations of Motion and Numerical Simulations

The equations of motion for the signal field,  $\hat{S}$ , and the anti-Stokes field,  $\hat{A}$ , interacting with an ensemble of atoms via a strong control pulse with Rabi frequency  $\Omega$ , and generating a spin-wave excitation  $\hat{B}$ , are given by [32]

$$(c\partial_z + \partial_t)\hat{S} = ic\sqrt{\frac{d\gamma}{L}\frac{\Omega}{\Gamma_s}}(\hat{B} + \hat{F}_s) - \kappa_s\hat{S},$$
 (A1)

$$(c\partial_z + \partial_t)\hat{A} = ic\sqrt{\frac{d\gamma}{L}}\frac{\Omega}{\Gamma_a}(\hat{B}^{\dagger} + \hat{F}_a) - \kappa_a \hat{A},$$
 (A2)

$$\partial_t \hat{B} = -i\Omega^* \sqrt{\frac{d\gamma}{L}} \left(\frac{1-\alpha}{\Gamma_s} + \frac{\alpha}{\Gamma_s^*}\right) \hat{S} + i\Omega^* \sqrt{\frac{d\gamma}{L}} \left(\frac{1-\alpha}{\Gamma_a} + \frac{\alpha}{\Gamma_a^*}\right) \hat{A}^{\dagger} - \left(\frac{1}{\Gamma_a^*} + \frac{1}{\Gamma_s}\right) |\Omega|^2 \hat{B} - \Omega^* \left(\frac{\hat{F}_a^{\dagger}}{\Gamma_a^*} + \frac{\hat{F}_s}{\Gamma_s}\right).$$
(A3)

Here,  $\Gamma_{s,a} = \gamma + i\Delta_{s,a}$  is the complex detuning of the signal and anti-Stokes fields, and  $\gamma = \gamma_N + \gamma_P$  is the total Lorentzian linewidth of the excited state including the natural linewidth  $\gamma_N$  and pressure broadening due to collisions with the buffer gas,  $\gamma_P$ . *d* is the pressure-broadened optical depth of the ensemble, which is related to the on-resonance optical depth  $d_0$  by  $d = d_0 \gamma_N / \gamma$ , and  $\alpha$  is the proportion of atoms that remains in state  $|3\rangle$  due to finite optical pumping efficiency. The atomic absorption and dispersion are described by

$$\kappa_s = \frac{d\gamma c}{L} \left[ \frac{(1-\alpha)}{\Gamma_s} + \frac{\alpha}{\Gamma_s^* + i\Delta_{\rm hf}} \right],\tag{A4}$$

$$\kappa_a = \frac{d\gamma c}{L} \left[ \frac{(1-\alpha)}{\Gamma_a + i\Delta_{\rm hf}} + \frac{\alpha}{\Gamma_a^*} \right],\tag{A5}$$

where  $\Delta_{hf}$  is the ground-state splitting between hyperfine levels  $|1\rangle$  and  $|3\rangle$ .  $\hat{F}_{s,a}$  are the Langevin noise operators which are introduced in addition to the decay terms on the atomic coherences, described by  $\Gamma_{s,a}$ , to account for fluctuations and ensure that the bosonic commutation relations still hold. These operators will only appear in normally ordered expectation values in expressions for the memory efficiency, and since the operators are initially in the vacuum state these expectation values are zero and we neglect such terms in our numerical simulations. However, if there is significant occupation of the excited state due to linear absorption of the control field, it may be that this assumption is no longer valid, and hence our model may not capture all sources of noise.

We solve Eqs. (A1)–(A3) numerically using a combination of the Runge-Kutta method and Chebyshev iteration method, and consider three cases:

(1) BNS-Raman. We consider a range of detunings around the absorption condition  $\Delta_s = -2\Delta_{\rm hf}$ .

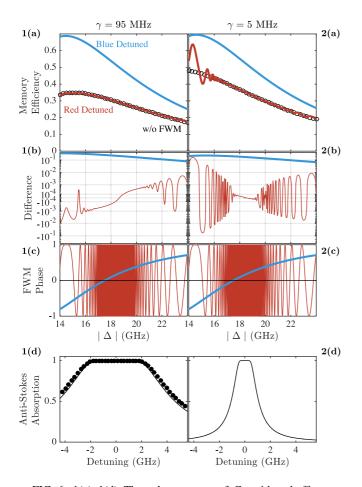


FIG. 6. 1(a)–1(d) The relevant case of Cs with a buffer gas,  $\gamma = 95$ MHz; 2(a)–2(d) without a buffer gas,  $\gamma = \gamma_N$ . Note that (a)–(c) have the absolute detuning  $|\Delta|$ . (a) Numerical simulations of the memory efficiency for three cases: STD-Raman [Blue Detuned: blue (upper) line], BNS-Raman [Red Detuned: red (lower) line], and ideal Raman without four-wave mixing in the equations (w/o FWM: black circles). (b) Difference between blue- and red-detuned memory efficiency and "w/o FWM" memory efficiency in (a). (c) Four-wavemixing phase for the blue- and red-detuned cases. (d) Anti-Stokes absorption. Note that the center of the feature is at zero detuning. For 1(d), we include the absorption measured with a narrowband laser (black dots).

(2) STD-Raman. A similar range of detunings as in (1) is used for the blue-detuned case  $\Delta_s = +2\Delta_{\rm hf}$ .

(3) Ideal Raman. We artificially turn off the four-wavemixing process and calculate the memory efficiency for the ideal, red-detuned Raman memory.

The results of these simulations for an on-resonance optical depth of  $d_0 = 2.9 \times 10^4$  and a control pulse energy of 750 pJ are shown in Fig. 6, both for a pressure-broadened linewidth of  $\gamma = 96$  MHz (left) and the natural linewidth  $\gamma = \gamma_N = 5.2$  MHz (right). We see that the efficiency in case (2) is higher than the ideal case (3) due to four-wave-mixing gain. The efficiency for the BNS-Raman [case (1)] is almost identical to the ideal case (3) over a broad range of detunings around the condition  $\Delta = -2\Delta_{\rm hf} = -18.4$  GHz due to the built-in noise suppression.

The second panel in Fig. 6 1(b) and 2(b) shows the difference in efficiency between cases (1) and (3) (red), and cases (2) and (3) (blue), on a logarithmic scale, and quantifies the increase in efficiency due to four-wave-mixing gain. We see that the four-wave-mixing gain is significantly suppressed in the BNS scheme over the whole 10 GHz frequency range, and at the exact absorption condition we see a suppression in the gain process by over four orders of magnitude compared to the standard Raman memory.

We note that the efficiency in case (1) oscillates around the ideal case (seen more clearly for the case with no pressure broadening) and this is due to the four-wave-mixing phase-matching condition. The phase mismatch,  $\delta k = 2k_c - k_s - k_a$ , is shown in the third panel in Fig. 6 1(c) and 2(c), and we see that it is rapidly changing due to the strong absorption feature at the anti-Stokes frequency. The four-wave-mixing process goes in and out of phase, resulting in an energy transfer back and forth between the four fields. We note that the dispersive feature is very broad, and the four-wave-mixing process is poorly phase matched over a wide range of detunings around  $\Delta = -2\Delta_{\rm hf}$ . This allows suppression of SFWM noise using atomic absorption even for a broadband noise field, and therefore this BNS scheme is widely applicable for narrow or broadband memory protocols.

# 2. Experimental Setup

To experimentally implement the Raman memory in warm caesium vapor, we use the  $6 S_{1/2} (F = 3)$  and (F = 4) hyperfine states as ground and storage states,  $|1\rangle$  and  $|3\rangle$ , respectively, which are separated by  $\Delta_{\rm hf} = 9.2$  GHz. The memory interaction is mediated by strong control pulses which drive an off-resonant two-photon Raman transition from the ground to storage state via the  $6P_{3/2}$  manifold,  $|2\rangle$ . The signal and control pulses are generated using a fiber-integrated electrooptic modulator to carve pulses from a continuous wave laser. The modulator features two electrodes which are driven by arbitrary wave-form generators with a sampling rate of 50 Gs/s and which facilitate arbitrary phase and amplitude control of the output pulses. We generate three pulses of an intensity full-width half maximum of 10 ns to act as the signal, read-in, and read-out control pulses. The signal pulse passes through a further bulk electro-optic modulator which is modulated at a frequency of 9.2 GHz to generate sidebands. The carrier frequency and blue sideband are filtered away using a Fabry-Pérot etalon, leaving the red sideband which is in twophoton resonance with the control field. The control pulses are amplified using a tapered amplifier to ensure sufficient pulse energy to drive the memory interaction. The orthogonally polarized signal and control pulses are temporally and spatially overlapped, and focused to a waist radius of 130  $\mu$ m in the center of a caesium vapor cell. The vapor is heated to a temperature of 83.0 °C to give a resonant optical depth of  $d_0 = 2.98 \times 10^4$ , and placed inside a  $\mu$ -metal magnetic shield to reduce magnetic dephasing of the spin wave. The system is initialized by optically pumping the ensemble into the F = 3ground state via a counterpropagating continuous-wave laser on resonance with the  $6S_{1/2} \rightarrow 6P_{1/2}$  transition. The EOM is switched off during the memory interaction using an EOM (EOSpace) to prevent depletion of the stored spin wave. A

buffer gas of 5 Torr of N<sub>2</sub> is mixed with the caesium vapor to allow a high pumping efficiency of  $(1 - \alpha) = 99.85\%$  to be reached at such high optical depths [33].

After the memory interaction, the strong control field is filtered away from the retrieved signal using a series of polarization and frequency filtering. First a calcite beam displacer suppresses the control field by over five orders of magnitude, before the signal is coupled into a single-mode fiber. A series of Fabry-Pérot etalons (FPEs) are then used to further suppress the control field: four FPEs with a free-spectral range (FSR) of 18.4 GHz to maximally suppress the control frequency (9.2 GHz separated from the signal), followed by two FPEs with an FSR of 103 GHz, to further suppress the control field in addition to suppressing the anti-Stokes field (which is 18.4 GHz separated from the signal and therefore transmits through the first FPEs) and the broadband fluorescence noise. In total, a suppression of the control field of over 110 dB is achieved, while the transmission of the signal field is 15%. After filtering, the signal is coupled into a single-mode fiber and detected with a standard fiber-coupled single-photon avalanche photodiode and a time-to-digital converter.

# 3. Derivation of $g_{out}^{(2)}$

The equations of motion for the Raman memory are given by Eqs. (A1)–(A3). These equations are linear so the resulting evolution may be characterized by Green's function mappings  $G_{ij}$  from initial mode *i* to final mode *j*, which may be represented in the form

$$\begin{pmatrix} \hat{S}_{\text{out}} \\ \hat{A}^{\dagger}_{\text{out}} \\ \hat{B}_{\text{out}} \end{pmatrix} = \begin{pmatrix} G_{ss} & G_{sa^{\dagger}} & G_{sb} \\ G_{a^{\dagger}s} & G_{a^{\dagger}a^{\dagger}} & G_{a^{\dagger}b} \\ G_{bs} & G_{ba^{\dagger}} & G_{bb} \end{pmatrix} \begin{pmatrix} \hat{S}_{\text{in}} \\ \hat{A}^{\dagger}_{\text{in}} \\ \hat{B}_{\text{in}} \end{pmatrix}.$$
 (A6)

We consider the input signal to be a superposition of Fock states:

$$|s_{\mathrm{in}}\rangle = \sum_{n} c_{n} |n\rangle, \quad N_{\mathrm{in}} \equiv \langle s_{\mathrm{in}} | S^{\dagger}S |s_{\mathrm{in}}\rangle.$$

We take the input anti-Stokes field to be vacuum,  $|a_{in}\rangle = |0_a\rangle$ , and the initial spin wave to have population  $\langle b_{in} \rangle = \alpha$ , given by number of unpumped atoms in initial state  $|3\rangle$ . The full input state is described as  $|\Psi_{in}\rangle = |s_{in}\rangle |0_a\rangle |b_{in}\rangle$  and the number of photons retrieved from the memory is given by

$$N_{\rm out} = \langle \Psi_{\rm in} | S_{\rm out}^{\dagger} S_{\rm out} | \Psi_{\rm in} \rangle \equiv N_{\rm mem} + N_{\rm SRS}^{AS} + N_{\rm SRS}^{P}, \quad (A7)$$

where the first term describes output photons due to the desired memory interaction, the second term describes output photons arising from spontaneous four-wave mixing, and the third term is spontaneous Raman scattering due to the initial occupation of the spin wave from imperfect optical pumping.

The second-order autocorrelation function of the input signal is

$$g_{\rm in}^{(2)} = \frac{\langle \psi_{\rm in} | S^{\dagger} S^{\dagger} S S | \psi_{\rm in} \rangle}{\langle \psi_{\rm in} | S^{\dagger} S | \psi_{\rm in} \rangle^2} = \frac{\langle \psi_{\rm in} | S^{\dagger} S^{\dagger} S S | \psi_{\rm in} \rangle}{\langle N_{\rm in} \rangle^2},$$

and so

$$\langle \psi_{
m in} | S^{\dagger}S^{\dagger}SS | \psi_{
m in} 
angle = \langle N_{
m in} 
angle^2 g_{
m in}^{(2)}$$

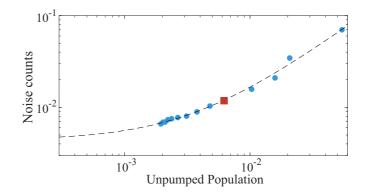


FIG. 7. Measured noise counts per pulse as a function of the proportion of the population in the storage state,  $\alpha$ . The red square indicates the parameters that were used for the data presented in the main text. The black dashed line is a linear fit to the data.

By evaluating expressions for all the nonzero terms in  $\langle S^{\dagger}S^{\dagger}SS \rangle$ , we find

$$\begin{split} \langle S^{\dagger}S^{\dagger}SS \rangle &= 2\langle S^{\dagger}S \rangle^{2} - N_{\mathrm{in}}^{2} \left( 2\eta^{2} - g_{\mathrm{in}}^{(2)}\mathcal{G}_{ss} \right) \\ &- 2 \left[ \left( N_{\mathrm{SRS}}^{AS} \right)^{2} - \mathcal{G}_{sa^{\dagger}} \right] - 2 \left[ \left( N_{\mathrm{SRS}}^{P} \right)^{2} - \alpha^{2}\mathcal{G}_{sb} \right], \end{split}$$

where  $\mathcal{G}_{ij} \equiv \iint dt dt' |G_{ij}(t,t')|^4$ . Therefore,

$$g_{\text{out}}^{(2)} = 2 - \left\{ N_{\text{in}}^2 \left( 2\eta^2 - g_{\text{in}}^{(2)} \mathcal{G}_{ss} \right) + 2 \left[ \left( N_{\text{SRS}}^{AS} \right)^2 - \mathcal{G}_{sa^{\dagger}} \right] \right. \\ \left. + 2 \left[ \left( N_{\text{SRS}}^P \right)^2 - \alpha^2 \mathcal{G}_{sb} \right] \right\} / \left( \eta N_{\text{in}} + N_{\text{SRS}}^{AS} + N_{\text{SRS}}^P \right)^2.$$
(A8)

We assume that the noise due to spontaneous Raman scattering would give thermal output statistics, i.e.,  $g_{out}^{(2)}(N_{in} = 0) = 2$ . We hence approximate  $\mathcal{G}_{sa^{\dagger}} = (N_{SRS}^{AS})^2$  and  $\mathcal{G}_{sb} = (N_{SRS}^p)^2/\alpha^2$ , which simplifies Eq. (A8) to

$$g_{\text{out,sig}}^{(2)} = 2 - \frac{N_{\text{in}}^2 \left(2\eta^2 - g_{\text{in}}^{(2)} \mathcal{G}_{ss}\right)}{(\eta N_{\text{in}} + N_{\text{SRS}})^2},$$
 (A9)

where  $N_{\text{SRS}} = N_{\text{SRS}}^{AS} + N_{\text{SRS}}^{P}$ .

We treat the noise contribution from fluorescence as an incoherent sum of fields with  $N_F$  photons and  $g_F^{(2)}$ . The incoherent sum of fields 1 and 2 is given by [22]

$$g_{12}^{(2)} = \frac{N_1^2 g_1^{(2)} + 2N_1 N_2 + N_2^2 g_2^{(2)}}{(N_1 + N_2)^2}$$

and hence

$$g_{\text{tot}}^{(2)} = 1 + \frac{N_s^2 - N_{\text{in}}^2 (2\eta^2 - g_{\text{in}}^{(2)} \mathcal{G}_{ss})}{N_{\text{tot}}^2} + \frac{N_F^2 (g_F^{(2)} - 1)}{N_{\text{tot}}^2},$$

where photons produced in the signal mode are  $N_{\rm S} = \eta N_{\rm in} + N_{\rm SRS}$ , and the total number of photons produced is  $N_{\rm tot} = \eta N_{\rm in} + N_{\rm SRS} + N_F$ . This leads to the expression for  $g_{\rm out}^{(2)}$  given in the main text:

$$g_{\text{out}}^{(2)} = 1 + \frac{aN_{\text{out}}^2 + 2N_{\text{SRS}}N_{\text{out}} + b}{(N_{\text{out}} + N_{\text{SRS}} + N_{\text{F}})^2},$$
 (A10)

where

$$a = g_{in}^{(2)} \mathcal{G}_{ss} / \eta^2 - 1,$$
  

$$b = N_{SRS}^2 + N_F^2 (g_F^{(2)} - 1).$$
(A11)

### 4. Noise vs Optical Pumping

To distinguish whether the remaining noise from spontaneous Raman scattering,  $N_{\text{SRS}}$ , is due to four-wave mixing or Stokes scattering due to imperfect optical pumping, we measure the average number of noise photons per pulse as a function of the pumping efficiency,  $1 - \alpha$ . The results, shown in Fig. 7, indicate that the noise decreases linearly with the amount of residual population in the storage state. The control pulse energy here is 330 pJ and the storage time is 50 ns. However, these data were taken with the optical pumping beam turned on all the time, and not switching off

- A. I. Lvovsky, B. C. Sanders, and W. Tittel, Nat. Photon. 3, 706 (2009).
- [2] M. D. Eisaman, A. André, F. Massou, M. Fleischhauer, A. S. Zibrov, and M. D. Lukin, Nature (London) 438, 837 (2005).
- [3] K. F. Reim, P. Michelberger, K. C. Lee, J. Nunn, N. K. Langford, and I. A. Walmsley, Phys. Rev. Lett. 107, 053603 (2011).
- [4] J. Wolters, G. Buser, A. Horsley, L. Béguin, A. Jöckel, J.-P. Jahn, R. J. Warburton, and P. Treutlein, Phys. Rev. Lett. 119, 060502 (2017).
- [5] M. Namazi, C. Kupchak, B. Jordaan, R. Shahrokhshahi, and E. Figueroa, Phys. Rev. Appl. 8, 034023 (2017).
- [6] M. Zugenmaier, K. B. Dideriksen, A. S. Sørensen, B. Albrecht, and E. S. Polzik, Commun. Phys. 1, 76 (2018).
- [7] K. T. Kaczmarek, P. M. Ledingham, B. Brecht, S. E. Thomas, G. S. Thekkadath, O. Lazo-Arjona, J. H. D. Munns, E. Poem, A. Feizpour, D. J. Saunders, J. Nunn, and I. A. Walmsley, Phys. Rev. A 97, 042316 (2018).
- [8] J. Guo, X. Feng, P. Yang, Z. Yu, L. Q. Chen, C.-H. Yuan, and W. Zhang, Nat. Commun. 10, 148 (2019).
- [9] Y. O. Dudin, L. Li, and A. Kuzmich, Phys. Rev. A 87, 031801(R) (2013).
- [10] D.-S. Ding, W. Zhang, Z.-Y. Zhou, S. Shi, G.-Y. Xiang, X.-S. Wang, Y.-K. Jiang, B.-S. Shi, and G.-C. Guo, Phys. Rev. Lett. 114, 050502 (2015).
- [11] Y.-W. Cho, G. T. Campbell, J. L. Everett, J. Bernu, D. B. Higginbottom, M. T. Cao, J. Geng, N. P. Robins, P. K. Lam, and B. C. Buchler, Optica 3, 100 (2016).
- [12] E. Saglamyurek, T. Hrushevskyi, A. Rastogi, K. Heshami, and L. J. LeBlanc, Nat. Photon. 12, 774 (2018).
- [13] Y.-F. Hsiao, P.-J. Tsai, H.-S. Chen, S.-X. Lin, C.-C. Hung, C.-H. Lee, Y.-H. Chen, Y.-F. Chen, I. A. Yu, and Y.-C. Chen, Phys. Rev. Lett. **120**, 183602 (2018).
- [14] P. Vernaz-Gris, K. Huang, M. Cao, A. S. Sheremet, and J. Laurat, Nat. Commun. 9, 363 (2018).
- [15] B. Lauritzen, J. Minár, H. de Riedmatten, M. Afzelius, N. Sangouard, C. Simon, and N. Gisin, Phys. Rev. Lett. 104, 080502 (2010).
- [16] M. P. Hedges, J. J. Longdell, Y. Li, and M. J. Sellars, Nature (London) 465 (2010).
- [17] G. Heinze, C. Hubrich, and T. Halfmann, Phys. Rev. Lett. 111, 033601 (2013).

for memory storage. Due to photorefractive damage of the switching EOM, we can only achieve a low pumping power and a poor extinction ratio in this current demonstration. The red dot in Fig. 7 indicates the conditions corresponding to the measurements for the autocorrelation data in the main text. By increasing the optical pumping power to 5 mW, we will be able to reduce the noise from  $N_{\text{noise}} = 11.8(2) \times 10^{-3}$  to  $N_{\text{noise}} = 6.6(2) \times 10^{-3}$ . If we extrapolate these data to perfect optical pumping efficiency,  $\alpha \rightarrow 0$ , we find a linear offset of  $N_{\text{noise}} = (4.4 \pm 1.7) \times 10^{-3}$ , which we attribute to fluorescence noise.

- [18] D. G. England, K. A. G. Fisher, J.-P. W. MacLean, P. J. Bustard, R. Lausten, K. J. Resch, and B. J. Sussman, Phys. Rev. Lett. 114, 053602 (2015).
- [19] A. Seri, A. Lenhard, D. Rieländer, M. Gündogan, P. M. Ledingham, M. Mazzera, and H. de Riedmatten, Phys. Rev. X 7, 021028 (2017).
- [20] K. Heshami, D. G. England, P. C. Humphreys, P. J. Bustard, V. M. Acosta, J. Nunn, and B. J. Sussman, J. Mod. Opt. 63, 2005 (2016).
- [21] M. Gündogan, P. M. Ledingham, K. Kutluer, M. Mazzera, and H. de Riedmatten, Phys. Rev. Lett. 114, 230501 (2015).
- [22] P. S. Michelberger, T. F. M. Champion, M. R. Sprague, K. T. Kaczmarek, M. Barbieri, X. M. Jin, D. G. England, W. S. Kolthammer, D. J. Saunders, J. Nunn, and I. A. Walmsley, New J. Phys. 17, 043006 (2015).
- [23] G. Romanov, C. O'Brien, and I. Novikova, J. Mod. Opt. 63, 2048 (2016).
- [24] N. Prajapati, G. Romanov, and I. Novikova, J. Opt. Soc. Am. B 34, 1994 (2017).
- [25] D. J. Saunders, J. H. D. Munns, T. F. M. Champion, C. Qiu, K. T. Kaczmarek, E. Poem, P. M. Ledingham, I. A. Walmsley, and J. Nunn, Phys. Rev. Lett. 116, 090501 (2016).
- [26] K. Zhang, J. Guo, L. Q. Chen, C. Yuan, Z. Y. Ou, and W. Zhang, Phys. Rev. A 90, 033823 (2014).
- [27] M. Dabrowski, R. Chrapkiewicz, and W. Wasilewski, Opt. Express 22, 26076 (2014).
- [28] V. Popkov, S. Essink, C. Presilla, and G. Schütz, Phys. Rev. A 98, 052110 (2018).
- [29] A. V. Gorshkov, A. André, M. D. Lukin, and A. S. Sørensen, Phys. Rev. A 76, 033805 (2007).
- [30] A. Christ, K. Laiho, A. Eckstein, K. N. Cassemiro, and C. Silberhorn, New J. Phys. 13, 033027 (2011).
- [31] F. Kaneda, K. Garay-Palmett, A. B. U'Ren, and P. G. Kwiat, Opt. Express 24, 10733 (2016).
- [32] J. Nunn, J. H. D. Munns, S. Thomas, K. T. Kaczmarek, C. Qiu, A. Feizpour, E. Poem, B. Brecht, D. J. Saunders, P. M. Ledingham, D. V. Reddy, M. G. Raymer, and I. A. Walmsley, Phys. Rev. A 96, 012338 (2017).
- [33] S. E. Thomas, J. H. D. Munns, K. T. Kaczmarek, C. Qiu, B. Brecht, A. Feizpour, P. M. Ledingham, I. A. Walmsley, J. Nunn, and D. J. Saunders, New J. Phys. **19**, 063034 (2017).