optica

On-demand indistinguishable single photons from an efficient and pure source based on a Rydberg ensemble: supplement

D. P. ORNELAS-HUERTA,[1] (iD) A. N. CRADDOCK,[1] (iD) E. A. GOLDSCHMIDT,[2,3] A. J. HACHTEL,[1] Y. WANG,[1] P. BIENIAS,[1,4] A. V. GORSHKOV,[1,4] S. L. ROLSTON,[1] AND J. V. PORTO[1,*]

[1] *Joint Quantum Institute, NIST/University of Maryland, College Park, Maryland 20742, USA*
[2] *Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green Street, Urbana, Illinois 61801, USA*
[3] *U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783, USA*
[4] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*
Corresponding author: porto@umd.edu

On-demand indistinguishable single photons from an efficient and pure source based on a Rydberg ensemble: supplementary material

D.P. Ornelas-Huerta[1], **A.N. Craddock**[1], **E.A. Goldschmidt**[2,3], **A.J. Hachtel**[1], **Y. Wang**[1], **P. Bienias**[1,4], **A.V. Gorshkov**[1,4], **S.L. Rolston**[1], and **J.V. Porto**[1,*]

[1] *Joint Quantum Institute, NIST/University of Maryland, College Park, MD 20742, USA*
[2] *Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green Street, Urbana, Illinois 61801, USA*
[3] *US Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783, USA*
[4] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, MD 20742, USA*
[*] *Corresponding author: porto@umd.edu*

Compiled July 6, 2020

Single photons coupled to atomic systems have shown to be a promising platform for developing quantum technologies. Yet a bright on-demand, highly pure, and highly indistinguishable single-photon source compatible with atomic platforms is lacking. In this work, we demonstrate such a source based on a strongly interacting Rydberg system. The large optical nonlinearities in a blockaded Rydberg ensemble convert coherent light into a single collective excitation that can be coherently retrieved as a quantum field. We simultaneously observe a fully single-mode (spectral, temporal, spatial, and polarization) efficiency up to 0.098(2), a detector-background-subtracted $g^{(2)} = 5.0(1.6) \times 10^{-4}$, and indistinguishability of 0.980(7), at an average photon production rate of $1.18(2) \times 10^4$ s^{-1}. All of these make this system promising for scalable quantum information applications. Furthermore, we investigate the effects of contaminant Rydberg excitations on the source efficiency, and observed single-mode efficiencies up to 0.18(2) for lower photon rates. Finally, recognizing that many quantum information protocols require a single photon in a fully single mode, we introduce metrics that take into account all degrees of freedom to benchmark the performance of on-demand sources.

1. DETAILED EXPERIMENTAL CONFIGURATION

All the experiments are carried out with $\approx 10^4$ ^{87}Rb atoms trapped in a three-beam crossed optical dipole trap with 1003-nm wavelength. Two of the beams form a $\approx \pm 11°$ with respect to the x-axis (along the probe direction), while a third elliptical shaped beam travels in the y-axis, with all beams in the same (x-y) plane. The relative powers of the dipole beams are adjusted so that the RMS dimensions of the trapped atomic cloud are $\sigma_r = 20$ μm in the radial direction and $\sigma_x = 27$ μm.

The initial trapping and cooling take place in a magneto-optical trap (MOT). For most experiments, we load for 250 ms; if

we need to adjust the atomic medium optical density (OD), we change the loading time, ranging from 50 ms to 1500 ms (with OD up to $\approx$16). Afterward, we perform a compressed-MOT stage by ramping-up the magnetic field gradient, while at the same time slowly ramping-up the dipole trap power.

We further cool the atoms to ≈ 10 μK using a gray molasses [**?**]. Next, we optically pump the atoms into the $|5S_{1/2}, F = 2, m_F = 2\rangle$ state, using σ_+ polarized light blue-detuned from the $F = 2$ to $F' = 2$, D1 transition. We then couple the ground and Rydberg state with a two-photon transition. A 780-nm weak-probe field addresses the transition from the ground state, $|5S_{1/2}, F = 2, m_F = 2\rangle$ to the intermedi-

ate state, $|5P_{3/2}, F = 3, m_F = 3\rangle$; a strong control field addresses the transition from the intermediate state to the Rydberg state, $|139S_{1/2}, J = 1/2, m_J = 1/2\rangle$ with a wavelength of 479 nm.

Both the probe and control lasers are frequency stabilized via an ultra-low-expansion (ULE) cavity with a linewidth < 10 kHz. We use probe light that has been transmitted and filtered by the ULE cavity to reduce phase noise [?] .

There are eight electrodes in vacuum that allow for control of local electric fields. With this configuration, we cancel DC Stark shifts to tens of kHz level in all three directions, shifts that would otherwise tune the Rydberg state out of resonance due to the large polarizability of the 139S state, $\alpha_{139S} \approx 61$ GHz/(V/cm)^2 [?].

The axial RMS of the atomic cloud, $\sigma_x \approx 27$ μm is smaller than the blockade radius, $r_b \approx 60$ μm to suppress the creation of multiple Rydberg atoms. Additionally, we focus the probe beam down to a $1/e^2$ waist of $w_p \approx 3.3$ μm to ensure the system is effectively uni-dimensional ($w_p \leq r_b$). The control beam is counter-propagating to the probe and focused to a beam waist of $w_c \approx 19$ μm. The larger beam waist provides an approximately uniform control field across the probe area. After exiting the chamber, the probe light passes through a polarization beamsplitter (PBS), and a set of bandpass filters centered at 780 nm, a narrow 1-nm bandwidth filter (Alluxa 780-1 OD6[1]), and a broader 12.5-nm bandwidth filter (Semrock LL01-780-12.5), before being coupled into a single-mode polarization-maintaining fiber (PMF). Then, the light is sent to a Hong-Ou-Mandel (HOM) interferometer, which has another set of broad filters in front of the single-mode fibers (SMF) that send the light to the single-photon avalanche detectors (SPAD) (Excelitas SPCM-780-13). The count events from the SPADs are recorded by an externally triggered time-tagging device (Roithner TTM8000) with a 82.3 ps resolution.

We write a spin wave by pulsing the probe and the control field for ≈ 370 ns. The peak Rabi frequencies are $\Omega_p/(2\pi) \approx 1$ MHz and $\Omega_c/(2\pi) \approx 7$ MHz, respectively. Both fields are detuned from the intermediate state by $\Delta_p/(2\pi) \approx 50$ MHz, with the two-photon transition close to resonance. Due to the collective nature from the blockaded excitation [?], there is a $\sqrt{N} \approx 20$ enhancement to the two-photon Rabi frequency, $\sqrt{N}\Omega_{2\text{-photon}} = \sqrt{N}\Omega_p\Omega_c/(2\Delta_p)$, inferred from the π-time. This enhancement corresponds to an OD≈ 13 given the blockaded volume.

After writing, we turn off the addressing lasers and hold (store) the spin wave in the medium for ≈ 350 ns; this is the minimum time required to switch the control acousto-optic modulator (AOM) frequency. We turn on the control field blue-detuned from the intermediate state by $\Delta_c/(2\pi) \approx 7$ MHz to map the spin wave into a single photon. We use an AOM before the PMF as a shutter to avoid saturating the SPADs during the initial write pulse.

We measure the optical losses along the path of the probe light to characterize the generation efficiency in Table S1. The propagation efficiency includes all the optical elements, such as filters, dichroics, mirrors, polarizing beamsplitters, mirrors, and lenses. With realistic improvements on higher transmission coatings and using an electro-optical modulator instead of an AOM to shutter the write pulse, we could get an efficiency up to 0.65 after the PMF, from the current 0.44.

Element	Efficiency
Optics transmission	0.75(2)
AOM diffraction	0.79(2)
PMF coupling	0.75(2)
HOM interferometer	0.38(1)
SPAD	0.67(1)

Table S1. List of the efficiencies along the probe path.

2. BACKGROUND SUBTRACTION

For all our single-photon measurements, we use two SPADs, with average background rates of ≈ 80 s^{-1} and ≈ 100 s^{-1}. This count rate is due to detector dark counts and leakage of ambient light.

Figure S1 shows the detected SPAD counts as a function of time during a pulse period t_p=5 μs, plotted on a log scale. In addition to the photon pulse, a small amount of leakage of the write pulse through the AOM shutter is evident. Since the photons arrive at the detectors at a known time, we apply a software gate corresponding to a 1.4 μs time window (indicated by the blue-dashed lines), which contains more than 99.9% of the pulse. The gate is implemented during the post-processing of the data and is equivalent to hardware gating the detectors. We determine the mean detector background singles count rate by measuring the counts during a time window (indicated by the red dashed lines) in which the shutter AOM is off. The background count rate when the AOM is on increases by $\simeq 4\%$ (not easily visible in Fig. S1), indicating a small amount of additional background produced by the apparatus. We do not include this additional background in the subtraction described below.
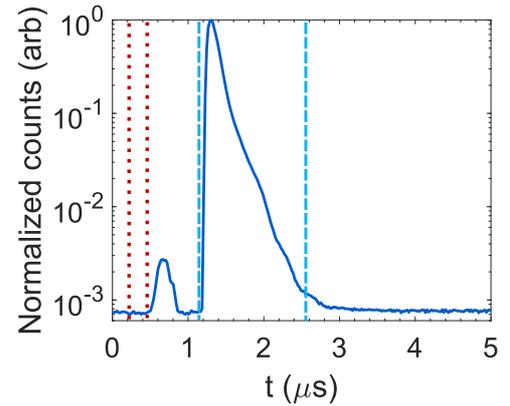
Fig. S1. Dark-blue solid line shows the counts recorded by SPAD 1 for a full cycle of 5 μs. The red-dotted lines show the window where we extract the mean background probability per unit time $B_1(B_2)$ when the shutter AOM is off. The light-blue dashed lines indicate the gating window to obtain the photon signal probability per unit time $P_1(t_1)$, $P_1(t_2)$ for SPAD 1, SPAD 2. The smaller peak at the left of the photon signal are leakage counts from the write pulse.

From the measured background noise level for each detector and the single photon profile, we determine the background-photon and background-background coincidences profile within

---

[1]The identification of commercial products in this paper does not imply recommendation or endorsement by the National Institute of Standards and Technology or the Army Research Laboratory, nor does it imply that the items identified are necessarily the best available for the purpose.
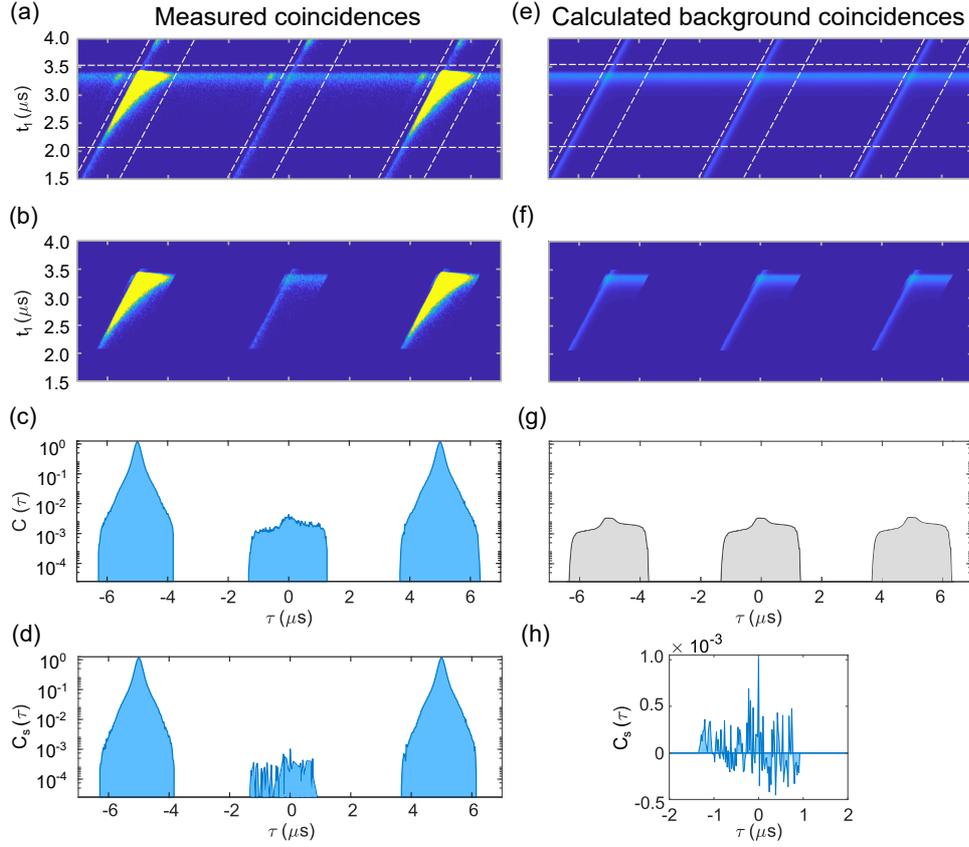
Fig. S2. Reconstruction of coincidences given the background and photon rate measured at each detector. (a) Raw data coincidences as a function of absolute time t_1, for SPAD 1 and τ, the relative time between both SPADs. White dashed lines indicate the position of the gating window for each repetition cycle. (b) Data with gate applied. (c) Total coincidence rate after applying the gate as a function of τ. (d) Photon-photon coincidence rate after subtracting the background from the data. (e) Calculated background coincidence $c_{\text{back}}(t_1, \tau)$, based on the measured single-event rates $P_i(t)$ and B_i. (f) Calculated background coincidences after the gate. (g) Gated-background coincidence rate $C_{\text{back}}(\tau)$. (h) Zoom around $\tau = 0$ of background-subtracted data in linear scale.

the gate window and subtract it.

The probability of a background coincidence, c_{back}, is the sum of the products of single event rates:

$$c_{\text{back}}(t_1, t_2) = P_1(t_1)B_2(t_2) + B_1(t_1)P_2(t_2) + B_1(t_1)B_2(t_2), \quad \text{(S1)}$$

where t_1 and t_2 are absolute times relative to a clock, for SPAD 1 and 2 respectively. $P_i(t_i)$, is the probability per unit time of a photon detection event at detector i, and $B_i(t_i)$ is the probability per unit time of a background count. Changing the relative time coordinate, $\tau = t_1 - t_2$, the background coincidence probability is

$$c_{\text{back}}(t_1, \tau) = P_1(t_1)B_2(\tau + t_1) + B_1(t_1)P_2(\tau + t_1) + B_1(t_1)B_2(\tau + t_1). \quad \text{(S2)}$$

We integrate t_1 over a time window $t_{\text{end}} - t_{\text{start}}$ to obtain the total background coincidence rate as a function of the relative time, τ:

$$C_{\text{back}}(\tau) = \int_{t_{\text{start}}}^{t_{\text{end}}} dt_1 \left[P_1(t_1)B_2(\tau + t_1) + B_1(t_1)P_2(\tau + t_1) + B_1(t_1)B_2(\tau + t_1) \right], \quad \text{(S3)}$$

where, t_{start}, is synchronized to the photon arrival. With the gate,

the background and pulse probability have a time dependence

$$B_1(t_1), P_1(t_1) = \begin{cases} B_1, P_1(t_1) & t_{\text{start}} \leq t_1 \leq t_{\text{end}} \\ 0 & \text{otherwise} \end{cases}$$

$$B_2(\tau + t_1), P_2(\tau + t_1) =$$
$$\begin{cases} B_2, P_2(\tau + t_1) & t_1 - t_{\text{end}} \leq \tau \leq t_1 - t_{\text{start}} \\ 0 & \text{otherwise} \end{cases}$$

This process is shown graphically in Figure S2, where C_{back} are the total coincidences rate from photon-background and background-background around $\tau = 0$. Finally Figure S2(h) shows the background subtracted coincidences rate, $C_s(\tau)$, within the gate window.

3. HOM VISIBILITY DISCUSSION

If two single photons are incident simultaneously on separate ports a_1 and a_2 of a perfect 50:50 beamsplitter (BS) the initial state $|1_1, 1_2\rangle$, becomes:

$$|1_1, 1_2\rangle \rightarrow \frac{1}{\sqrt{2}}\left(|2_3, 0_4\rangle + |0_3, 2_4\rangle\right) \quad \text{(S4)}$$

where a_3, a_4 are the output ports and we assumed that the input photons are in pure states and indistinguishable from each other. In this case, the probability of a coincidence detection is zero and the HOM visibility is one. In practice, the following factors reduce the visibility from its maximum value [?]:

- one or both photons are not in a pure state,

- there is more than one photon at either BS input port,

- an imperfect 50:50 BS.

We will focus on the effect of the last two conditions: multiphoton events and imperfect BS.

Following the discussion from [?], we define the scattering matrix, S for a general BS as,

$$S = \begin{pmatrix} t_1 & r_2 e^{i\phi_2} \\ r_1 e^{i\phi_1} & t_2 \end{pmatrix}, \qquad (S5)$$

where r_1 (r_2), t_1 (t_2), are the reflection and transmission amplitudes with a relative phase ϕ_1 (ϕ_2) for port 1 (2).

Then the input-output relations of the BS, ignoring any frequency dependence:

$$\begin{pmatrix} \hat{a}_3 \\ \hat{a}_4 \end{pmatrix} = \begin{pmatrix} t_1 & r_2 e^{i\phi_2} \\ r_1 e^{i\phi_1} & t_2 \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix}, \qquad (S6)$$

where $\hat{a}_i$ are the photon ladder operator for the input and output ports. Generally, the scattering matrix, S, is not unitary.

For a lossy BS, where the output fields' total energy is lower than the input fields' energy, the following inequality holds:

$$\sqrt{t_1^2 r_2^2 + r_1^2 t_2^2 + 2t_1 r_1 t_2 r_2 \cos \alpha} \le \sqrt{(1 - t_1^2 - r_1^2)(1 - t_2^2 - r_2^2)}, \qquad (S7)$$

where $\alpha = \phi_1 + \phi_2$, affects the maximum value that the visibility can attain. The phase, α, is constrained by energy conservation, and we assume $\alpha = \pi$.

The number operator for the input ports 1 and 2 (output 3 and 4) is $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$. Assuming that the probability of states with more than two photons is negligible and that the probability of a single photon P_1, and the probability of two photons P_2 are the same at both input ports (1 and 2), the incoming state:

$$\hat{\rho}_{\text{in}} = P_0 |0_1, 0_2\rangle \langle 0_1, 0_2| + P_1 |1_1, 0_2\rangle \langle 1_1, 0_2| + P_1 |0_1, 1_2\rangle \langle 0_1, 1_2|$$
$$+ P_1^2 |1_1, 1_2\rangle \langle 1_1, 1_2| + P_2 |2_1, 0_2\rangle \langle 2_1, 0_2| + P_2 |0_1, 2_2\rangle \langle 0_1, 2_2|, \qquad (S8)$$

The coincidence probability $P(1_3, 1_4)$ for $\hat{\rho}_{\text{in}}$,

$$P(1_3, 1_4) \approx \text{Tr}\{\hat{\rho}_{\text{in}} \hat{n}_3 \hat{n}_4\}$$
$$\approx \text{Tr}\{\hat{\rho}_{\text{in}}[t_1^2 r_1^2 \hat{n}_1(\hat{n}_1 - 1) + t_2^2 r_2^2 \hat{n}_2(\hat{n}_2 - 1)$$
$$+ (t_1^2 t_2^2 + r_1^2 r_2^2 - 2t_1 r_1 t_2 r_2)\hat{n}_1 \hat{n}_2]\}$$
$$= (t_1^2 r_1^2 + t_2^2 r_2^2)2P_2 + (t_1^2 t_2^2 + r_1^2 r_2^2 - 2ct_1 r_1 t_2 r_2)P_1^2, \qquad (S9)$$

where c is the mode overlap in all degrees of freedom of the two incident photons, and it is present in the term where the photons exit the same output port of the BS. Here, the mode overlap is the parameter that describes the indistinguishability obtained from a HOM-interference measurement, and can its value ranges from zero to one. Following the assumption that the probability of states with more than two photons is negligible, we can rewrite

Port/Polarization	T	R
Port 1 H	0.502(5)	0.421(3)
Port 1 V	0.484(5)	0.428(3)
Port 2 H	0.511(9)	0.426(5)

Table S2. Transmission and reflection coefficients for the BS used in the HOM interferometer.

P_2 as a function of the correlation function $g^{(2)}(0)$ and P_1, as $P_2 \approx g^{(2)}(0)P_1^2/2$. The coincidence probability:

$$P(1_3, 1_4) \approx \left[t_1^2 t_2^2 + r_1^2 r_2^2 + (t_1^2 r_1^2 + t_2^2 r_2^2)g^{(2)} - 2ct_1 r_1 t_2 r_2\right] P_1^2. \qquad (S10)$$

For the more general case, where the BS coefficients are not the same for orthogonal polarizations, H, and V

$$\mathcal{V} = \frac{P(1_3, 1_4)_{HV}\big|_{c=0} - P(1_3, 1_4)_{HH}\big|_{c=c}}{P(1_3, 1_4)_{HV}\big|_{c=0}}$$
$$= \left[t_{1_V}^2 t_{2_H}^2 + r_{1_V}^2 r_{2_H}^2 - t_{1_H}^2 t_{2_H}^2 - r_{1_H}^2 r_{2_H}^2 \right.$$
$$\left. + (t_{1_V}^2 r_{1_V}^2 - t_{1_H}^2 r_{1_H}^2)g^{(2)} + 2ct_{1_H} r_{1_H} t_{2_H} r_{2_H}\right]$$
$$\times \left[t_{1_V}^2 t_{2_H}^2 + r_{1_V}^2 r_{2_H}^2 + (t_{1_V}^2 r_{1_V}^2 + t_{2_H}^2 r_{2_H}^2)g^{(2)}\right]^{-1}, \qquad (S11)$$

where we assume that in the case of $P(1_3, 1_4)_{HV}$, the photon at port 1 has H-polarization and the photon at port 2 has V-polarization, similarly for $P(1_3, 1_4)_{HH}$, both incoming photons have H-polarization.

In the particular case of a BS with symmetric ports, $t_1^2 = t_2^2 = T$ and, $r_1^2 = r_2^2 = R$, the visibility reduces to:

$$\mathcal{V} = \frac{2c}{T/R + R/T + 2g^{(2)}}. \qquad (S12)$$

If $T = R = 1/2$ and $g^{(2)}(0) = 0$, then the visibility is equal to the incoming photons overlap, c.

In Table S2, we show the measured transmission and reflection coefficients of the BS used in the HOM characterization, for both H- and V-polarization.

We measured a background-subtracted visibility to be $\mathcal{V} = 0.964(6)$, and using Eq. S11 to take into account the imperfect BS, we find a mode overlap of 0.980(7).

4. CONTAMINANTS

We use a simple model to characterize the effects of the contaminants on the photon generation, where there is a probability that a stored spin wave is converted to a contaminant. Once a contaminant is present in the medium, it disables the writing and storing of a spin wave until the contaminant decays, with a time constant τ_c. If P_c is the probability of creating a contaminant on a given pulse, then the probability, P_n, of a contaminant being present at pulse n depends on whether one was created in one of the previous pulses and remained to the n-th pulse

$$P_n = P_{n-1} e^{-t_p/\tau_c} + (1 - P_{n-1})P_c, \qquad (S13)$$

where t_p is the pulse spacing. If we set the initial condition to be $P_1 = P_c$, and use the identity, $(1 - x) \sum_{j=0}^{n-1} x^j = 1 - x^n$, we get the expression:
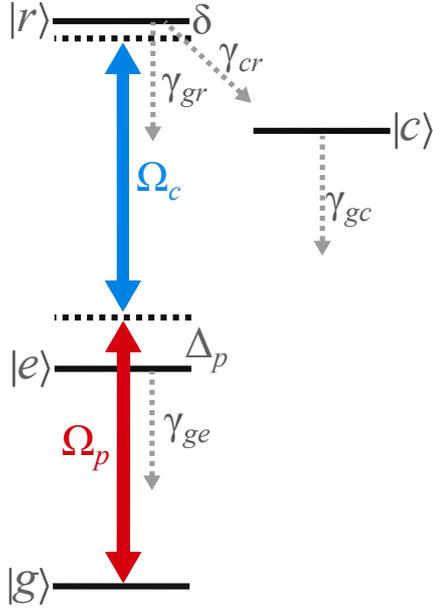
Fig. S3. Atomic levels showing the driving fields and decay rates used in theoretical model. Here we show the ground-state, as $|g\rangle = |5S_{1/2}\rangle$, the intermediate state $|e\rangle = |5P_{3/2}\rangle$, Rydberg state $|r\rangle = |139S_{1/2}\rangle$, and the contaminants states as $|c\rangle$.

$$P_n = P_c \frac{1 - (e^{-t_p/\tau_c} - P_c)^n}{1 - e^{-t_p/\tau_c} + P_c}. \tag{S14}$$

Then, the probability of successfully generating a photon, $P_g(n)$ is

$$
\begin{aligned}
P_g(n) &= P_{max}(1 - P_n) \\
&= P_{max}\left(1 - P_c \frac{1 - (e^{-t_p/\tau_c} - P_c)^n}{1 - e^{-t_p/\tau_c} + P_c}\right),
\end{aligned}
\tag{S15}
$$

where P_{max} is the maximum probability of generating a photon. For $n \to \infty$, the steady state probability P_s,

$$P_s \approx P_{max}\left(1 - \frac{P_c}{1 - e^{-t_p/\tau_c} + P_c}\right). \tag{S16}$$

We also model how the correlation function, $g^{(2)}(m\, t_p)$ for integer $m \neq 0$, is modified due to contaminants:

$$
\begin{aligned}
g^{(2)}(|m|t_p) &= \frac{\langle P_s P_g(m)\rangle}{\langle P_s^2\rangle} \\
&= 1 + P_c \frac{(e^{-t_p/\tau_c} - P_c)^m}{1 - e^{-t_p/\tau_c}}.
\end{aligned}
\tag{S17}
$$

This manifests as a bunching feature around $\tau = 0$.

5. THEORETICAL MODEL

A. Write and storage efficiency

For the writing and storage time, we model the spin wave as a super-atom with N atoms being collectively driven into a single excitation to the Rydberg state. The energy levels and decay rates of the super-atom are shown in Figure S3.

We simulated the writing stage as driving the super-atom from the ground to the Rydberg state, with $\sqrt{N}$-enhanced Rabi

frequency. During the writing time, t_w, the Rabi frequencies, $\Omega_p/(2\pi) = 1.0(2)$ MHz and $\Omega_c/(2\pi) = 6.8(3)$ MHz are kept constant. For the storage time, t_s, these driving frequencies are set to zero.

The Hamiltonian describing the the system depicted in Fig. S3 in the rotating wave approximation is given by:

$$H(t) = \frac{\hbar}{2}\begin{pmatrix} 0 & \sqrt{N}\Omega_p(t) & 0 & 0 \\ \sqrt{N}\Omega_p(t) & -2\Delta_p & \Omega_c(t) & 0 \\ 0 & \Omega_c(t) & -2\delta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{S18}$$

in the basis of $|g\rangle$, $|e\rangle$, $|r\rangle$, $|c\rangle$, for the ground, intermediate, Rydberg and contaminant state, respectively.

Using the Python package QuTip [?], we calculated the non-unitary dynamics of this first stage using the master equation for the four level density matrix ρ:

$$\dot{\rho} = -\frac{i}{\hbar}[H, \rho] + \sum_n \left(C_n \rho C_n^\dagger - \frac{1}{2}\{\rho, C_n^\dagger C_n\}\right), \tag{S19}$$

where $C_1 = \sqrt{\gamma_{ge}}\,|g\rangle\langle e|$, $C_2 = \sqrt{\gamma_{gr}}\,|g\rangle\langle r|$, $C_3 = \sqrt{\gamma_{cr}}\,|c\rangle\langle r|$, and $C_4 = \sqrt{\gamma_{gc}}\,|g\rangle\langle c|$ are the jump operators.

Given the decay rates of the different states: $\gamma_{ge}/(2\pi) = 6.9(6)$ MHz, $\gamma_{ge}/(2\pi) = 88(6)$ kHz, $\gamma_{rc}/(2\pi) = 5(1)$ kHz, and $\gamma_{gc}/(2\pi) = 2.5(3)$ kHz, we calculate that the writing and storage efficiency are $\eta_w = 0.82(1)$, $\eta_s = 0.82(1)$, respectively.

B. Retrieval efficiency

We follow the derivations in Ref. [?] to compute the retrieval efficiency. In the rescaled unit-less coordinates, $\tilde{z} = 0$ and $\tilde{z} = 1$ represent the front and the end of the atomic cloud, respectively. Suppose all atoms are in the $|r\rangle$ state in the beginning of the retrieval stage at time $\tilde{t} = 0$, the shape of the spin wave is given by $S(\tilde{z}, \tilde{t} = 0) = 1$ for $\tilde{z} \in [0,1]$ and $S(\tilde{z}, \tilde{t} = 0) = 0$ for $\tilde{z}$ elsewhere. The retrieval efficiency can be expressed in terms of the photon field $\mathcal{E}(\tilde{z}, \tilde{t})$ emitted by the stored spin wave at the end of the atomic cloud:

$$\eta_r = \int_0^\infty d\tilde{t}\,|\mathcal{E}(\tilde{z} = 1, \tilde{t})|^2. \tag{S20}$$

$\mathcal{E}(1, \tilde{t})$ can be calculated as:

$$
\begin{aligned}
\mathcal{E}(1, \tilde{t}) = &-\sqrt{d}\tilde{\Omega}(\tilde{t})\exp(-\tilde{\gamma}_s\tilde{t})\int_0^1 d\tilde{z}\,\frac{\exp\left(-\frac{h(0,\tilde{t})+d\tilde{z}}{1+i\tilde{\Delta}}\right)}{1+i\tilde{\Delta}} \\
&\times I_0\left(2\frac{\sqrt{h(0,\tilde{t})d\tilde{z}}}{1+i\tilde{\Delta}}\right) S(1 - \tilde{z}),
\end{aligned}
\tag{S21}
$$

where we define dimensionless parameters $d = \text{OD}/2$, $\tilde{\gamma}_s = (\gamma_{gr} + \gamma_{cr})/\gamma_{ge}$, $\tilde{\Delta} = 2\Delta_p/\gamma_{ge}$, $\tilde{\Omega}(t) = \Omega_c(t)/\gamma_{ge}$. $h(\tilde{t}, \tilde{t}') = \int_{\tilde{t}}^{\tilde{t}'} |\tilde{\Omega}(\tilde{t}'')|^2 d\tilde{t}''$ and I_0 is the 0th-order modified Bessel function of the first kind. When the control field Ω_c is constant in time, we define the dimensionless parameter $x_s = 2\tilde{\gamma}_s/|\tilde{\Omega}_c|^2$ which characterizes the strength of the decay rate compared to the control field. Eq. (S20) can be evaluated as

$$\eta_r = \int_0^1 d\tilde{z}\int_0^1 d\tilde{z}'\,K_r S(1 - \tilde{z})S^*(1 - \tilde{z}'), \tag{S22}$$

where K_r is given by

$$
\begin{aligned}
K_r = &\frac{df(x_s)}{2} \\
&\times \exp\left[-\frac{df(x_s)}{2}\left((1 + x_s(1 - i\tilde{\Delta}))\bar{z} + (1 + x_s(1 + i\tilde{\Delta}))\bar{z}'\right)\right] \\
&\times I_0(d\sqrt{\bar{z}\bar{z}'}f(x_s)),
\end{aligned}
$$

(S23)

and $f(x_s) = \frac{2}{2 + x_s(1 + \tilde{\Delta}^2)}$.

Evaluating the integral in Eq. (S22) numerically, we obtain the retrieval efficiency $\eta_r = 0.63(2)$. With these results, we estimate that the photon generation probability at the end of the cloud is $P_{\text{th}} = 0.42(3)$.

C. Possible improvements

With conservative feasible experimental improvements, such as implementing a ground-state blue-detuned optical dipole trap, as well as increasing the following parameters: $\Omega_c/(2\pi) \approx$ 10 MHz, $\Delta_p/(2\pi) \approx 100$ MHz and OD=20, while decreasing the spin wave dephasing by a factor of two, we estimate that we could increase our probabilities up to $\eta_w\eta_s = 0.86$ and $\eta_r = 0.72$, while maintaining a relatively low contaminant probability, $P_c \approx 3 \times 10^{-2}$.

From the theoretical model, the main limiting factor is the retrieval process; in principle, the retrieval efficiency increases with higher OD; however, the contaminant production also grows with OD. A Rydberg ensemble with low OD coupled to a cavity could further increase light-matter interactions and therefore increase the overall photon production probability, making it a promising platform for scalable quantum information applications.

6. SINGLE-PHOTON SOURCES

In Tables S3 and S4, there is detailed information about the properties of a representative sample of single-photon sources plotted in Fig. 5. in the main text. The notation, R, repetition rate, P is the fibered efficiency, V, is the indistinguishability, P_1^{sm} is the fully single-mode single photon probability, $\mathcal{R}$ is the fully single-mode single photon rate, and $\mathcal{F}$ is the fidelity.

Type	Ref	R (MHz)	P	V	$g^{(2)}$	P_1^{sm}	$\mathcal{R} \times 10^6\ (s^{-1})$	$\mathcal{F}$
SPDC	[?]	76	≈0.01	0.91	0.09	0.009	0.69	0.910
SPDC	[?]	80	0.049	0.962	0.03	0.047	3.74	0.961
MUX	[?]	10	≈0.002	0.91	∼0.2	0.002	0.02	0.910
MUX	[?]	0.5	0.667	0.91	0.269	0.562	0.28	0.843
QD	[?]	82	≈0.02	0.996	0.0028	0.020	1.63	0.996
QD	[?]	80	≈ 0.08	0.7	0.013	0.056	4.47	0.699
QD	[?]	76	0.337	0.93	0.027	0.312	23.71	0.926
QD	[?]	76	0.10	0.94	0.006	0.094	7.14	0.940
QD	[?]	76	0.24	0.975	0.025	0.233	17.77	0.972
QD	[?]	81	0.11	0.92	0.05	0.101	8.17	0.918

Table S3. Table comparing the performances of solid state single-photon sources: spontaneous parametric down conversion (SPDC), multiplexed-heralded-single-photon source (MUX) and, quantum dot (QD). Values estimated from available data.

Type	Ref	Duty Cycle (%)	R (MHz)	P	V	$g^{(2)}$	P_1^{sm}	$\mathcal{R} \times 10^3\ (s^{-1})$	$\mathcal{F}$
Yb ion	[?]	80	8	0.003	0.86	$\sim 10^{-3}$	0.003	18.16	0.860
Rb Atom	[?]	33	0.052	0.003	0.9	$\sim 10^{-3}$	0.003	0.05	0.899
Ensemble in cavity	[?]	≈ 1.8	0.05	0.08	0.9	0.05	0.072	0.06	0.898
Atom in cavity	[?]	≈ 2	0.7	0.2	0.7	$\sim 10^{-2}$	0.140	1.96	0.699
Atom in cavity	[?]	0.1	1	0.21	0.87	0.02	0.182	0.18	0.868
Atom in cavity	[?]	100	0.01	0.39	0.64	0.02	0.249	2.49	0.637
Rydberg	this work	60	0.2	0.100	0.982	5×10^{-4}	0.098	11.8	0.982
Rydberg	future	60	0.5	0.4	0.99	$\approx 10^{-4}$	≈ 0.4	120	0.99

Table S4. Table comparing the performances of different atomic single-photon sources. Here $\mathcal{R}$ is weighted by the duty cycle of operation. Values estimated from available data.