Supplemental Material for "Photon Subtraction by Many-Body Decoherence"

C. R. Murray[1], I. Mirgorodskiy[2], C. Tresp[3], C. Braun[3], A. Paris-Mandoki[3], A. V. Gorshkov[4], S. Hofferberth[3] and T. Pohl[1]

[1] *Center for Quantum Optics and Quantum Matter, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK 8000 Aarhus C, Denmark*

[2] *5. Phys. Inst. and Center for Integrated Quantum Science and Technology, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany*

[3] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, 5230 Odense-M, Denmark and*

4 *Joint Quantum Institute and Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*

I. GATE EXCITATION DENSITY MATRIX DYNAMICS

Here, we will outline the solution to the gate excitation density matrix given in Eq. (2) of the main text. To first describe the EIT dynamics of the source field, we introduce the bosonic operator $\hat{\mathcal{E}}^\dagger(z,t)$, which creates a source photon at position z and time t, and similarly introduce the operators $\hat{P}^\dagger(z,t)$, $\hat{S}^\dagger(z,t)$ and $\hat{C}^\dagger(z,t)$ which create collective atomic excitations in $|p\rangle$, $|s\rangle$ and $|c\rangle$ respectively. In a one-dimensional approximation, these operators are governed by the following Heisenberg equations of motion,

$$\partial_t\hat{\mathcal{E}}(z,t) = -c\partial_z\hat{\mathcal{E}}(z,t) + iG\hat{P}(z,t), \tag{S1}$$

$$\partial_t\hat{P}(z,t) = iG\hat{\mathcal{E}}(z,t) + i\Omega_s\hat{S}(z,t) - \gamma\hat{P}(z,t) + \hat{F}(z,t), \tag{S2}$$

$$\partial_t\hat{S}(z,t) = i\Omega_s\hat{P}(z,t) - i\int_0^L dz' V_{z,z'}\hat{C}^\dagger(z',t)\hat{C}(z',t)\hat{S}(z,t), \tag{S3}$$

$$\partial_t\hat{C}(z,t) = -i\int_0^L dz' V_{z,z'}\hat{S}^\dagger(z')\hat{S}(z')\hat{C}(z,t). \tag{S4}$$

Here, c is the vaccum speed of light, $G = g\sqrt{\rho_a}$ is the collectively enhanced coupling of the $|g\rangle - |p\rangle$ transition (where g is the single atom coupling and ρ_a is the homogenous atomic density), Ω_s is the Rabi frequency of the classical crontrol field driving the $|p\rangle - |s\rangle$ Rydberg transition, and γ is the decay rate of the intermediate state $|p\rangle$. We assume low-intensity source and gate fields such the $|s\rangle - |s\rangle$ and $|c\rangle - |c\rangle$ interactions can be neglected. The operator $\hat{F}(z,t)$ describes Langevin noise associated with the decay of the intermediate state [S1].

Considering a system of n_g stored gate excitions and n_s incident source photons, we introduce $|\Psi_{n_g,n_s}\rangle$ as the initial state. In the Heisenberg picture, this can be constructed explicitly as,

$$|\Psi_{n_g,n_s}\rangle = \frac{1}{\sqrt{n_g!n_s!}}\left[\frac{1}{\sqrt{c}}\int_{-\infty}^{\infty} dz h(-z/c)\hat{\mathcal{E}}^\dagger(z,0)\right]^{n_s} \times \left[\int_0^L d\vec{z}_{n_g}\mathcal{C}(\vec{z}_{n_g})\hat{C}^\dagger(z_1,0)\dots\hat{C}^\dagger(z_{n_g},0)\right]|0\rangle, \tag{S5}$$

where $h(t)$ is the temporal mode of the incident (uncorrelated) source field, and $\mathcal{C}(\vec{z}_{n_g})$ is the initial spatial mode of the stored gate excitations where $\vec{z}_{n_g} \equiv z_1, z_2, \cdots, z_{n_g}$ denotes the vector of gate excitation coordinates. To determine the scattering-induced spin wave decoherence, it is necessary to consider the density matrix dynamics of the stored gate excitations. For this, we first define the operator $\hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)$

$$\hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = \prod_{i=1}^{n_g}\hat{C}^\dagger(x_i,t)\prod_{i=1}^{n_g}\hat{C}(y_i,t). \tag{S6}$$

This can then be used in conjuction with Eq. (S5) to define the elements of the stored spin wave density matrix as

$$\rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = \langle\Psi_{n_g,n_s}|\hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)|\Psi_{n_g,n_s}\rangle, \tag{S7}$$

which characterises the spatial coherence between different configurations $\vec{x}_{n_g}$ and $\vec{y}_{n_g}$ of the stored gate excitations in response to scattering n_s source photons. To evaluate the time dynamics of $\rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)$, we begin with the equation of motion for the coherence operator,

$$\partial_t \hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = i \int_0^L dz \left[\sum_k V_{z,x_k} - \sum_k V_{z,y_k} \right] \hat{S}^\dagger(z,t) \hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) \hat{S}(z,t), \tag{S8}$$

which can be readily derived from Eq. (S4). The solution to the spin wave operator $\hat{S}(z,t)$ will be a convolution of the form

$$\hat{S}(z,t) = \int_{-\infty}^{\infty} dt' \hat{e}(z, t-t') \hat{\mathcal{E}}(0,t'), \tag{S9}$$

where $\hat{e}(z,t)$ is an operator object which is intrinsically nonlinear in the stored spin wave density $\hat{C}^\dagger(z,t)\hat{C}(z,t)$. The general solution also includes terms propotional to $\hat{\mathcal{E}}(z,0)$, $\hat{P}(z,0)$, $\hat{S}(z,0)$ and $\hat{F}(z,0)$. However, since all our results only involve normally ordered expectation values, such terms give vanishing contributions for the initial state in Eq. (S5) [S2, S3]. With the definition for $\hat{S}(z,t)$ in Eq. (S9), the equation of motion for $\rho_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)$ can then be written as,

$$\begin{aligned} \partial_t \rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = i \frac{n_s}{c} \int_0^L dz \left[\sum_k V_{z,x_k} - \sum_k V_{z,y_k} \right] \int_{-\infty}^{\infty} dt' h^*(t') \int_{-\infty}^{\infty} dt'' h(t'') \\ \times \langle \Psi_{ng,n_s-1} | \hat{e}^\dagger(z, t-t') \hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) \hat{e}(z, t-t'') | \Psi_{ng,n_s-1} \rangle, \end{aligned} \tag{S10}$$

where we have used the property $\hat{\mathcal{E}}(0,t)|\Psi_{ng,n_s}\rangle = \hat{\mathcal{E}}(-ct,0)|\Psi_{ng,n_s}\rangle = \sqrt{n_s/c} h(t) |\Psi_{ng,n_s-1}\rangle$. In the limit where the source field is narrowband in relation to the EIT bandwidth, we can make the replacement $\hat{e}(z,t) = \hat{e}(z)\delta(t)$, where $\hat{e}(z)$ defines the static solution to $\hat{S}(z,t)$ as $\hat{S}(z, t\to\infty) = \hat{e}(z)\hat{\mathcal{E}}(0, t\to\infty)$. This can be obtained by solving Eqs. (S1 - S3) in the steady state to yield

$$\hat{e}(z) = -\frac{G}{\Omega_s} \frac{1}{1 + i \int dz' \mathcal{V}_{z,z'} \hat{C}^\dagger(z') \hat{C}(z')} \exp\left(\frac{d_b}{z_b} \int_0^z dz' \left[\frac{1}{1 + i \int dz'' \mathcal{V}_{z',z''} \hat{C}^\dagger(z'') \hat{C}(z'')} - 1 \right] \right), \tag{S11}$$

where $\mathcal{V}_{z,z'} = \gamma V_{z,z'}/\Omega_s^2$ is the rescaled interaction potential, and $2d_b = 2G^2 z_b/c\gamma$ is the optical depth per blockade radius, where z_b is defined according to $V_{z_b,0} = \Omega_s^2/\gamma$. Eq. (S10) can then be written as,

$$\partial_t \rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = i \frac{n_s}{c} |h(t)|^2 \int_0^L dz \left[\sum_k V_{z,x_k} - \sum_k V_{z,y_k} \right] \langle \Psi_{ng,n_s-1} | \hat{e}^\dagger(z) \hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) \hat{e}(z) | \Psi_{ng,n_s-1} \rangle. \tag{S12}$$

To proceed, we note that since the operator $\hat{e}(z)$ is constructed from the local density operator $\hat{C}^\dagger(z)\hat{C}(z)$, it conserves the total number of gate excitations. As such, the state $|C(\vec{x}_{n_g})\rangle = \prod_{i=1}^{n_g} \hat{C}^\dagger(x_i)|0\rangle$ is an eigenstate of $\hat{e}^\dagger(z)$ with an eigenvalue $e^*(z, \vec{x}_{n_g})$ defined by $\hat{e}^\dagger(z)|C(\vec{x}_{n_g})\rangle = e^*(z,\vec{x}_{n_g})|C(\vec{x}_{n_g})\rangle$, which can be readily derived from Eq. (S11) as

$$e(z, \vec{x}_{n_g}) = -\frac{G}{\Omega_s} \frac{1}{1 + i \sum_k \mathcal{V}_{z,x_k}} \exp\left(\frac{d_b}{z_b} \int_0^z dz' \left[\frac{\sum_k \mathcal{V}_{z,x_k}}{i - \sum_k \mathcal{V}_{z,x_k}} \right] \right). \tag{S13}$$

Upon then redefining $\hat{\rho}(\vec{x}_{n_g}, \vec{y}_{n_g}) = |C(\vec{x}_{n_g})\rangle\langle C(\vec{y}_{n_g})|$, it follows that the equation of motion for $\rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)$ can be written as

$$\partial_t \rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t) = n_s \phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}) \rho_{n_s-1}(\vec{x}_{n_g}, \vec{y}_{n_g}, t), \tag{S14}$$

where

$$\phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}) = i \frac{d_b}{z_b} \frac{\Omega_s^2}{G^2} \int_0^L dz \left[\sum_k \mathcal{V}_{z,x_k} - \sum_k \mathcal{V}_{z,y_k} \right] e^*(z, \vec{x}_{n_g}) e(z, \vec{y}_{n_g}), \tag{S15}$$

which defines the expression in Eq. (3) of the main text as $\Phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}) = 1 + \phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g})$. The system of equations for $\rho_{n_s}(\vec{x}_{n_g}, \vec{y}_{n_g}, t)$ goverened by Eq. (S14) can then be solved recusively in n_s to yield the final expression for the many-body density matrix given in Eq. (2) of the main text.

II. SPIN WAVE DECOHERENCE IN THE INFINITE d_b LIMIT

Here, we will derive the simple expression for the many-body density matrix in the infinite d_b limit. Upon spatially ordering all gate excitations, whereby x_1, y_1 are the coherence coordinates of the first excitation, x_2, y_2 are the coordinates of the second and so on, then the result for $\Phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g})$ can be approximated as

$$\Phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}) \approx 1 + \phi_1(x_1, y_1) + (1 - p_{<2})\phi_1(x_2, y_2) + (1 - p_{<3})\phi_1(x_3, y_3) + \cdots + (1 - p_{<n_g})\phi_1(x_{n_g}, y_{n_g}), \quad \text{(S16)}$$

where $\phi_1(x_k, y_k)$ is given by Eq. (S15), and $p_{<k}$ is the probability that a given source photon scatters before it reaches the k^{th} excitation. Here, it is implicitly assumed that $p_{<k}$ is close to unity, and in the infinite d_b limit, one can make the approximation $p_{<k} = 1$. In this case, $\Phi_{n_g}(\vec{x}_{n_g}, \vec{y}_{n_g}) \approx 1 + \phi_1(x_1, y_1) = \Phi_1(x_{\text{min}}, y_{\text{min}})$ as given by Eq. (4) of the main text.

III. APPROXIMATE MODEL OF RETRIEVAL EFFICIENCY

Here, we will derive the approximate model of retrieval efficiency presented in Eq. (5) of the main text. We start by considering a system of n_g stored gate excitations, and n_s photons in the incident source field. We assume a dilute system of excitations, such that the contributions from configurations of excitations with overlapping blockde radii can be neglected. The storage of such configurations will anyways be supressed due to self-blockade between gate photons. As a second simplification, we assume that the scattering induced localisation of one gate excitaiton does not affect the mode shape, and thus retrieval, of any other. Formally, this approximation can be implemented by assuming the gate photons are stored in non-overlapping modes, and we introduce $\rho_0^{(k)}(x_k, y_k)$ as the initial single body density matrix of the k^{th} excitation. With this simplification, the initial many-body density matrix is given by the pure (uncorrelated) state $\rho_0(\vec{x}_{n_g}, \vec{y}_{n_g}) = \rho_0^{(1)}(x_1, y_1)\rho_0^{(2)}(x_2, y_2)\cdots\rho_0^{(n_g)}(x_{n_g}, y_{n_g})$. The efficiency of retrieving the k^{th} excitation after source photon scattering can be calculated from its reduced density matrix $\rho_{n_s}^{(k)}(x, y)$, which can be calculated from the full many-body density matrix according to

$$\rho_{n_s}^{(k)}(x, y) = n_g \int dr_1 \cdots dr_{k-1} dr_{k+1} \cdots dr_{n_g} \rho_{n_s}(r_1, \cdots, r_{k-1}, x, r_{k+1}, \cdots, r_{n_g}, r_1, \cdots, r_{k-1}, y, r_{k+1}, \cdots, r_{n_g}). \quad \text{(S17)}$$

Assuming that the medium is much longer than the stored spin wave mode, the explicit form of $\rho_{n_s}^{(k)}(x, y)$ is given by,

$$\rho_{n_s}^{(k)}(x, y) = \left[1 + A^{k-1}\phi(x, y)\right]^{n_s} \rho_0^{(k)}(x, y), \quad \text{(S18)}$$

where $1 - A$ is the scattering probability per gate excitation defined according to,

$$A = \exp\left(\frac{d_b}{z_b}\int_{-\infty}^{\infty} dz' \left[\frac{\mathcal{V}_{z',0}}{i - \mathcal{V}_{z',0}} - \frac{\mathcal{V}_{z',0}}{i + \mathcal{V}_{z',0}}\right]\right), \quad \text{(S19)}$$

$$= \exp\left(2d_b \text{Re}\left[\frac{2\pi}{3}(-1)^{11/12}\right]\right), \quad \text{(S20)}$$

$$\approx \exp(-4d_b). \quad \text{(S21)}$$

The retrieval efficiency of the k^{th} excitation is then calculated as $\eta_k(n_s) = \mathcal{R}\left[\rho_{n_s}^{(k)}(x, y)\right]$. Here, $\mathcal{R}$ is a generic linear function for determining the retrieval efficiency from any given one-body density matrix and pulse sequence, whose explicit form is detailed in Ref. [S4]. To simplify the calculation of $\eta_k(n_s)$, we assume that the blockade radius is much smaller than the spatial extent of each spin wave mode. In this situation, photon scattering will practically cause complete localisation of a given stored gate excitation. The quantity $\phi(x, y)$ characterising this decoherence in Eq. (S18) can then be approximated by

$$\phi(x, y) = \begin{cases} 0 & \text{if } x = y \\ A - 1 & \text{otherwise} \end{cases} \quad \text{(S22)}$$

However, since the retrieval efficiency is predominatly determined by the spin wave coherences, it suffices to neglect the narrow digonal feature in $\phi(x, y)$ when caluclatuing $\eta_k(n_s)$. Using the approximation $\phi(x, y) \approx A - 1$, the retrieval efficiency of the k^{th} excitation is then given by

$$\eta_k(n_s) = \left[1 - p(1 - p)^{k-1}\right]^{n_s} \mathcal{R}\left[\rho_0^{(k)}(x, y)\right], \quad \text{(S23)}$$

where we have used the fact that the scattering probability per gate excitation is given by $p = 1 - A$. We can then calculate the total number of retrieved gate photons from the stored n_g-excitation Fock state after scattering n_s source photons as

$$\bar{n}_g(n_g, n_s) = \sum_{k=1}^{n_g} \eta_k(n_s) = \eta_R \sum_{k=1}^{n_g} \left[1 - p(1-p)^{k-1}\right]^{n_s}, \tag{S24}$$

where we have made use of the fact that the retrieval function is linear, and further assumed that the retrieval efficiency in the absence of photon scattering is the same for all gate excitations, i.e., $\mathcal{R}\left[\rho_0^{(k)}(x,y)\right] = \eta_R$. Finally, taking into account the coherent state nature of the involved fields, we can calculate the average number of retrieved gate photons by performing a coherent state average of $\bar{n}_g(n_g, n_s)$ over the number distribution of the gate and source fields, which ultimately yields

$$\bar{\alpha}_g = e^{-\alpha_g} e^{-\alpha_s} \sum_{n_g=1}^{\infty} \sum_{n_s=0}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} \frac{(\alpha_s)^{n_s}}{n_s!} \bar{n}_g(n_g, n_s), \tag{S25}$$

$$= \eta_R e^{-\alpha_g} e^{-\alpha_s} \sum_{n_g=1}^{\infty} \sum_{n_s=0}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} \frac{(\alpha_s)^{n_s}}{n_s!} \sum_{k=1}^{n_g} \left[1 - p(1-p)^{k-1}\right]^{n_s}, \tag{S26}$$

$$= \eta_R e^{-\alpha_g} \sum_{n_g=1}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} \sum_{k=1}^{n_g} \exp\left[-\alpha_s p(1-p)^{k-1}\right]. \tag{S27}$$

Finally, we can calculate the retrieval efficiency as the ratio of the number of retrieved gate photons with and without source field scattering,

$$\eta = \eta_R \frac{e^{-\alpha_g}}{\alpha_g} \sum_{n_g=1}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} \sum_{k=1}^{n_g} \exp\left[-\alpha_s p(1-p)^{k-1}\right], \tag{S28}$$

as given by Eq. (5) of the main text.

IV. SINGLE PHOTON SUBTRACTION VIA DECOHERENCE

Here we will derive a simple estimate for the efficiency of single photon subtraction based on the described decoherence mechanism. For this, first consider the operation using Fock states of the incoming gate and source fields. Let $|n_g\rangle$ describe the gate field containing n_g photons, and $|n_s\rangle$ describe the source field containing n_s photons. Through the combination of gate storage, source field scattering and gate retrieval, a perfectly functioning single photon subtractor will achieve the mapping $|n_g\rangle \mapsto |n_g - 1\rangle$. Taking into account a finite storage and retrieval efficiency due to linear losses, this photon subtraction can be achieved either from failed storage or failed retrieval, the latter of which is controlled via scattering induced decoherence.

To calculate the overall success probability for this to occur, let us first consider the storage losses. For this, we assume that storage is a linear process, and that each gate photon is stored with an probability η_S. The probability that all n_g photons are succesfully stored, $P_0^{(S)}(n_g)$, and the probability that one fails to store, $P_1^{(S)}(n_g)$, are then each given by

$$P_0^{(S)}(n_g) = \eta_S^{n_g} \tag{S29}$$

$$P_1^{(S)}(n_g) = n_g(1-\eta_S)\eta_S^{n_g-1} \tag{S30}$$

Assuming that $\bar{n}_g$ photons are stored, we then need to consider the subsequent decoherence dynamics from source field scattering. The probability $\mathrm{p}_0(\bar{n}_g)$ that an incoming source photon fails to scatter from any of the $\bar{n}_g$ stored gate excitations is given by

$$\mathrm{p}_0(\bar{n}_g) = (1-p)^{\bar{n}_g}, \tag{S31}$$

and the probability $\mathrm{p}_1(k_g)$ that a source photon scatters from the k_g^{th} excitation is given by,

$$\mathrm{p}_1(k_g) = p(1-p)^{k_g-1}, \tag{S32}$$

where p is the scattering probability per gate excitation. The probability $P_0^{(D)}(\bar{n}_g, n_s)$ that none of the n_s incoming source photons are scattered, such that $\bar{n}_g$ coherent excitations remain after the source field propagation, is then simply given by,

$$P_0^{(D)}(\bar{n}_g, n_s) = [\mathrm{p}_0(\bar{n}_g)]^{n_s} \tag{S33}$$

We then need to consider the probability that one gate photon is decohered after the source field scattering, which therefore leaves $\bar{n}_g - 1$ retrievable gate excitations. For this, the probability that n_s incoming source photons decohere the k_g^{th} gate excitation only can then be considered as a sum of contributions: either all n_s source photons scatter off the k_g^{th} excitation, or $n_s - 1$ source photons scatter off the k_g^{th} excitation while one is transmitted, or $n_s - 2$ source photons scatter off the k_g^{th} excitation while two are transmitted, and so on. The individual probabilities, $\mathcal{P}_{\bar{n}_g, n_s}^{(k_g, k_s)}$, that k_s out of the n_s incoming source photons scatter off the k_g^{th} gate excitation are then given by,

$$\mathcal{P}_{\bar{n}_g, n_s}^{(k_g, k_s)} = \binom{n_s}{k_s} [\mathrm{p}_1(k_g)]^{k_s} [\mathrm{p}_0(\bar{n}_g)]^{n_s - k_s}, \tag{S34}$$

where the binomial coefficient takes into account all the relevant scattering possibilities. The probability that at least one source photon scatters off the k_g^{th} gate excitation is then given by $\sum_{k_s=1}^{n_s} \mathcal{P}_{\bar{n}_g, n_s}^{(k_g, k_s)}$, such that the probability that only one gate excitation is left decohered after the passage of n_s source photons is given by

$$P_1^{(D)}(\bar{n}_g, n_s) = \sum_{k_g=1}^{\bar{n}_g} \sum_{k_s=1}^{n_s} \mathcal{P}_{\bar{n}_g, n_s}^{(k_g, k_s)} \tag{S35}$$

$$= \sum_{k_g=1}^{\bar{n}_g} \left\{ \sum_{k_s=0}^{n_s} \binom{n_s}{k_s} [\mathrm{p}_1(k_g)]^{k_s} [\mathrm{p}_0(\bar{n}_g)]^{n_s - k_s} - [\mathrm{p}_0(\bar{n}_g)]^{n_s} \right\} \tag{S36}$$

$$= \sum_{k_g=1}^{\bar{n}_g} \left\{ [\mathrm{p}_1(k_g) + \mathrm{p}_0(\bar{n}_g)]^{n_s} - [\mathrm{p}_0(\bar{n}_g)]^{n_s} \right\} \tag{S37}$$

We finally need to describe the linear retrieval losses, where we account for a finite retrieval probability of η_R per gate excitation. Assuming that we are left with $\tilde{n}_g$ coherent gate excitations after source field scattering, the probability that all are successfully retrieved, $P_0^{(R)}(\tilde{n}_g)$, and the probability that one is lost during retrieval, $P_1^{(R)}(\tilde{n}_g)$, are then each given by

$$P_0^{(R)}(\tilde{n}_g) = \eta_R^{\tilde{n}_g} \tag{S38}$$

$$P_1^{(R)}(\tilde{n}_g) = \tilde{n}_g (1 - \eta_R) \eta_R^{\tilde{n}_g - 1} \tag{S39}$$

The overall success probability for single photon subtraction $P_1(n_g, n_s)$ can than be evaluated by summing all contributions where exactly one photon is removed either during storage or retrieval,

$$\begin{aligned} P_1(n_g, n_s) = \; & P_1^{(S)}(n_g) P_0^{(D)}(n_g - 1, n_s) P_0^{(R)}(n_g - 1) \\ & + P_0^{(S)}(n_g) P_1^{(D)}(n_g, n_s) P_0^{(R)}(n_g - 1) \\ & + P_0^{(S)}(n_g) P_0^{(D)}(n_g, n_s) P_1^{(R)}(n_g) \end{aligned} \tag{S40}$$

The first line corresponds to single photon loss during storage, followed by perfect retrieval of all remaining stored excitations. The second line corresponds to successful storage of all gate photons, while one is removed upon retrieval due to scattering induced decoherence. Finally, the third line corresponds to successful storage of all gate photons, while one is removed upon retrieval due to linear losses. We can then use this result to obtain the success probability $P_1(n_g, \alpha_s)$ for single photon subtraction using a coherent source field containing an average number of photons α_s by performing a coherent state average of $P_1(n_g, n_s)$ over the number distribution of the source field,

$$P_1(n_g, \alpha_s) = e^{-\alpha_s} \sum_{n_s=0}^{\infty} \frac{\alpha_s^{n_s}}{n_s!} P_1(n_g, n_s) \tag{S41}$$

which is valid for $n_g > 0$. At this point, we can examine the effects of imperfect storage and retrieval. For a given n_g, we can find the source field intensity $\alpha_s^{(\text{opt})}$ that optimises $P_1(n_g, \alpha_s)$ under conditions of perfect storage and retrieval,

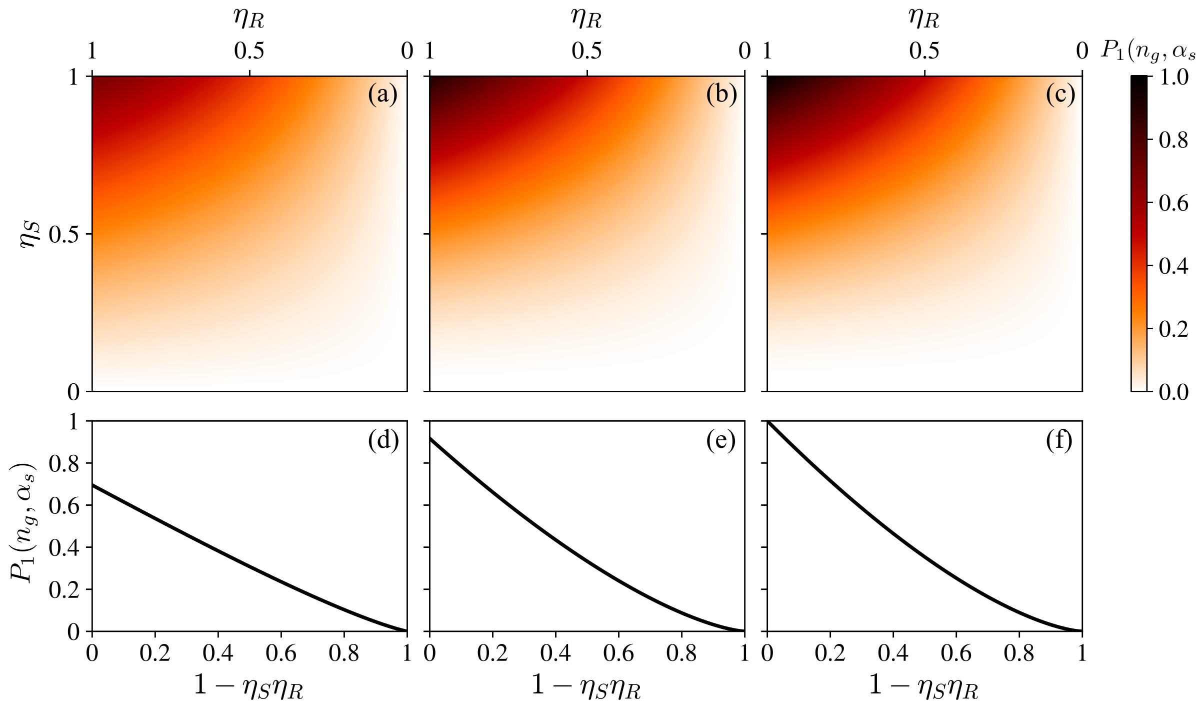

FIG. S1. (a-c) The subtraction efficiency $P_1(n_g, \alpha_s)$ for a Fock state with $n_g = 2$ incident gate photons is plotted as a function of the efficiency of storage, η_S, and retrieval, η_R. The blockaded optical depth is $d_b = 0.5, 1$ and 5 in (a), (b) and (c) respectively, and in each figure, we fix the coherent source field intensity to $\alpha_s^{(\mathrm{opt})}$ which optimises $P_1(n_g, \alpha_s)$ for perfect storage and retrieval, $\eta_S = \eta_R = 1$. $P_1(n_g, \alpha_s)$ is plotted as a function of the combined effieincy for storage and retrieval $\eta_S\eta_R$ (specifically for $\eta_S = \eta_R$) for $d_b = 0.5, 1$ and 5 in (d), (e) and (f) respectively.

$\eta_S = \eta_R = 1$. Considering a two-photon Fock state, we plot $P_1(n_g, \alpha_s^{(\mathrm{opt})})$ against η_S and η_R in Fig. S1, and further examine its scaling with the combined effieiency for storage and retrieval, $\eta_S\eta_R$.

Finally, considering a coherent state of the gate field, we can define the averaged single photon subtraction efficiency $\mathcal{F}$ defined in Eq. 6 of the main text by performing a coherent state average over the number distribution of the gate field,

$$\mathcal{F} = e^{-\alpha_g}\left[1 + \sum_{n_g=1}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} P_1(n_g, \alpha_s)\right]. \tag{S42}$$

Note that we implicitly set $P_1(n_g = 0, \alpha_s) = 1$, which assumes the subtraction is perfect for the vacuum component of the gate field. By optimisng $\mathcal{F}$ with respect to α_s for a given α_g, we obtain the blue curve in Fig. 4 of the main text (where we consider perfect storage and retrieval efficiency).

V. SINGLE PHOTON SUBTRACTION VIA SATURABLE ABSORPTION

Here, we will discuss the subtraction efficiency of the single photon absorber using a free-space Rydberg superatom, as recently demonstrated in [S5]. The general mechanism in this case relies on saturating the absorption of an optically thick ensemble via Rydberg blockade. Here, engineered dephasing on the Rydberg state with a rate Γ is used to achieve incoherent photon storage with a probability p. By working with a medium that is shorter than the blockade volume, the produced single Rydberg excitation then prevents any further photon absorption. For a large single photon detuning, the remaining off-resonant two-level medium is largely transparent to all subsequent photons, which scatter with a small residual probability $\tilde{p}$. Efficient single photon absorption with this mechanism then requires a large absorption probability p, while simultaneous minimising the residual photon losses.

This scheme is realised by coupling the quantised gate field to the low-lying excited state $|p\rangle$ with a large single photon detuning Δ. A continuously applied control field then couples $|p\rangle$ to the Rydberg state $|s\rangle$ on two-photon resonance with a Rabi frequency Ω. As before, one can introduce the operator $\hat{\mathcal{E}}^\dagger(z,t)$ to describe the creation of a gate photon, and introduce $\hat{P}^\dagger(z,t)$ and $\hat{S}^\dagger(z,t)$ to describe the creation of collective atomic excitations in $|p\rangle$ and $|s\rangle$. For a single incoming photon, the system dynamics are characterised by the following equations,

$$\partial_t \hat{\mathcal{E}}(z,t) = -c\partial_z \hat{\mathcal{E}}(z,t) + iG\hat{P}(z,t), \tag{S43}$$

$$\partial_t \hat{P}(z,t) = iG\hat{\mathcal{E}}(z,t) + i\Omega\hat{S}(z,t) - [i\Delta + \gamma]\hat{P}(z,t), \tag{S44}$$

$$\partial_t \hat{S}(z,t) = i\Omega\hat{P}(z,t) - \Gamma\hat{S}(z,t). \tag{S45}$$

Here, Langevin noise can be neglected for the reasons outlined in Sec. I. To zeroth order in the photon bandwidth, this system of equations reduces to a single propagation equation for $\hat{\mathcal{E}}(z)$ as

$$\partial_z \hat{\mathcal{E}}(z) = -\frac{1}{l_{\text{abs}}} \frac{1}{\frac{\Gamma_{\text{EIT}}}{\Gamma} + 1 + i\frac{\Delta}{\gamma}} \hat{\mathcal{E}}(z) \tag{S46}$$

where $\Gamma_{\text{EIT}} = \Omega^2/\gamma$ is the resonant EIT bandwidth. For a medium of length z_b, the transmitted photon operator can be solved as

$$\hat{\mathcal{E}}(z_b) = \exp\left[-d_b \frac{1}{\frac{\Gamma_{\text{EIT}}}{\Gamma} + 1 + i\frac{\Delta}{\gamma}}\right] \hat{\mathcal{E}}(0) = \sqrt{1-p}e^{i\theta}\hat{\mathcal{E}}(0), \tag{S47}$$

where θ is the phase of the transmitted field, and p is the absorption probability, the latter of which is given by

$$p = 1 - \exp\left[-2d_b \frac{1 + \frac{\Gamma_{\text{EIT}}}{\Gamma}}{\left(1 + \frac{\Gamma_{\text{EIT}}}{\Gamma}\right)^2 + \left(\frac{\Delta}{\gamma}\right)^2}\right]. \tag{S48}$$

The residual (dissipative) scattering probability of the blockaded ensemble after photon absorption can then be straightforwardly obtained from the above expression by setting $\Gamma_{\text{EIT}} = 0$ to give

$$\tilde{p} = 1 - \exp\left[-2d_b \frac{1}{1 + \left(\frac{\Delta}{\gamma}\right)^2}\right]. \tag{S49}$$

To analyse the subtraction efficiency, first consider a Fock state of the incoming gate field containing n_g photons. Treating the photons sequentially, the probability that exactly one is absorbed into the medium, whilst all others are transmitted can be calculated as

$$P_1(n_g) = \sum_{k=1}^{n_g} p(1-p)^{k-1}(1-\tilde{p})^{n_g-k}. \tag{S50}$$

The subtraction efficiency for a coherent state with an average number of α_g photons is then given by

$$\mathcal{F} = e^{-\alpha_g}\left[1 + \sum_{n_g=1}^{\infty} \frac{(\alpha_g)^{n_g}}{n_g!} P_1(n_g)\right]. \tag{S51}$$

For a given d_b and α_g, the optimal subtraction efficiency can be determined from Eq. (S51) by maximising $\mathcal{F}$ with respect to Δ/γ and $\Gamma_{\text{EIT}}/\Gamma$ to obtain the red dashed curve in Fig. 4 of the main text. Here, the additional constraint $\Gamma_{\text{EIT}}/\Gamma \gg 1$ is imposed to ensure that incoherent photon absorption dominates over the dissipative scattering.

[S1] M. O. Scully and M. S. Zubairy, *Quantum optics* (Cambridge ; New York : Cambridge University Press, 1997).
[S2] A. V. Gorshkov, R. Nath, and T. Pohl, Phys. Rev. Lett. **110**, 153601 (2013).
[S3] C. R. Murray, A. R. Gorshkov, and T. Pohl, New J. Phys. **18**, 92001 (2016).
[S4] A. V. Gorshkov, A. André, M. Fleischhauer, A. S. Sørensen, and M. D. Lukin, Phys. Rev. Lett. **98**, 123601 (2007).
[S5] C. Tresp, C. Zimmer, I. Mirgorodskiy, H. Gorniaczyk, A. Paris-Mandoki, and S. Hofferberth, Phys. Rev. Lett. **117**, 223001 (2016).