Supplementary Materials

Cavity-enhanced Raman emission from a single color center in a solid

Shuo Sun, Jingyuan Linda Zhang, Kevin A. Fischer, Michael J. Burek, Constantin Dory, Konstantinos G. Lagoudakis, Yan-Kai Tzeng, Marina Radulaski, Yousif Kelaita, Amir Safavi-Naeini, Zhi-Xun Shen, Nicholas A. Melosh, Steven Chu, Marko Lončar, and Jelena Vučković

1. Derivation of assumptions required for single-excitation regime

In the discussion for Fig. 1 of the main text, we assume the system contains at most one excitation, so that we can truncate the infinite Jaynes-Cummings ladders to the level structure shown in Fig. 1(b). In the absence of phonon-mediated ground state relaxation, this assumption is always valid. To illustrate why this is the case, we plot the expanded Jaynes-Cummings ladders in Fig. S1(a). We have also denoted all the coherent and incoherent couplings of the system in this figure. Clearly, if the system is initially in the state $|g_1,0\rangle$, it is not possible to drive the system to the upper ladders formed by $|g_1,1\rangle$, $|g_2,2\rangle$ and $|e,2\rangle$ because there are no couplings that drive a lower-ladder state to the higher ladder. The single-excitation regime is fundamentally guaranteed because we use a laser to drive the transition $|g_1\rangle \leftrightarrow |e\rangle$ which is not coupled to the cavity.

When including the ground state relaxation from $|g_2\rangle$ to $|g_1\rangle$, it is possible to break the single excitation regime. Figure S1(b) illustrates the same expanded energy level structure as shown in Fig. S1(a), but we have included the decay channel from the state $|g_2,n\rangle$ to $|g_1,n\rangle$, where $n$ is the number of photons in the cavity. If the ground state relaxation rate $\gamma_{flip}$ is larger
than the cavity energy decay rate $\kappa$, then it is possible for the state $|g_2,1\rangle$ to jump to the state $|g_2,2\rangle$ before emitting a photon, by first decaying to the state $|g_1,1\rangle$ with a rate $\gamma_{\text{flip}}$ and then rotating to the state $|g_2,2\rangle$ under the coherent drive $\Omega$ and cavity coupling $g$. Therefore when accounting for ground state relaxation, we need the condition $\gamma_{\text{flip}} <<< \kappa$ to be satisfied in order to truncate the basis to states with only a single excitation.

**Figure S1.** System level structure with the expanded Jaynes-Cummings ladders. The panel (a) and (b) show the cases when we ignore and take into account the ground state relaxation respectively.
2. Derivation of effective Rabi frequency $\Omega_{\text{eff}}$

In this section, we derive the effective Rabi frequency between the state $|g_{1},0\rangle$ and the state $|g_{2},1\rangle$ after adiabatic elimination of the state $|e,0\rangle$. We focus on the truncated level structure shown in Fig. 1(b) of the main text, and consider only the coherent interactions of the system. These interactions are governed by the Hamiltonian given by

$$\hat{H} = \hbar \Delta |e,0\rangle \langle e,0| + \hbar \frac{\Omega}{2} (|e,0\rangle \langle g_{1},0| + h.c.) + \hbar g (|e,0\rangle \langle g_{2},1| + h.c.).$$  

(S1)

The equation of motion for the system is governed by the Schrodinger equation, given by

$$\frac{d}{dt} |\psi(t)\rangle = -\frac{i}{\hbar} \hat{H} |\psi(t)\rangle,$$

where $|\psi(t)\rangle = c_{1}(t)|g_{1},0\rangle + c_{2}(t)|g_{2},1\rangle + c_{3}(t)|e,0\rangle$ is the state of the system at time $t$. The coefficients $c_{1}(t)$, $c_{2}(t)$, and $c_{3}(t)$ evolves according to

$$\frac{d}{dt} c_{1}(t) = -i \frac{\Omega}{2} c_{3}(t),$$  

(S2)

$$\frac{d}{dt} c_{2}(t) = -igc_{3}(t),$$  

(S3)

$$\frac{d}{dt} c_{3}(t) = -i\Delta c_{3}(t) - i \frac{\Omega}{2} c_{1}(t) - igc_{2}(t).$$  

(S4)

In the limit $\Omega,g \ll \Delta$, the coefficient $c_{3}(t)$ varies adiabatically with small amplitude such that $|c_{3}(t)| \ll 1$ and $\frac{d}{dt} c_{3}(t) \approx 0$ (see Ref. [35] for a rigorous mathematical proof). Then, the evolution of the ground states $|g_{1},0\rangle$ and $|g_{2},1\rangle$ is independent of the excited state $|e,0\rangle$.

This allows solving Eq. S4 for

$$c_{3}(t) = -\frac{\Omega}{2\Delta} c_{1}(t) - \frac{g}{\Delta} c_{2}(t).$$  

(S5)

Substituting Eq. (S5) back into Eqs. (S2) and (S3), we obtain that

$$\frac{d}{dt} c_{1}(t) = i\frac{\Omega^{2}}{4\Delta} c_{1}(t) + i\frac{\Omega g}{2\Delta} c_{2}(t),$$  

(S6)
Therefore, we can treat the system as an effective two-level system comprising the states $|g_1, 0\rangle$ and $|g_2, 1\rangle$, with effective Hamiltonian

$$\hat{H}_{\text{eff}} = -\hbar \frac{\Omega^2}{4\Delta} |g_1, 0\rangle \langle g_1, 0| - \hbar \frac{g^2}{\Delta} |g_2, 1\rangle \langle g_2, 1| - \hbar \frac{\Omega_{\text{eff}}}{2} (|g_1, 0\rangle \langle g_2, 0| + \text{h.c.}), \quad (S8)$$

where $\Omega_{\text{eff}} = \frac{\Omega g}{\Delta}$ is the effective Rabi frequency.

### 3. Derivation of Raman emission rate

We first calculate the Raman emission rate from a $\Lambda$-system without a cavity. Figure S2(a) shows the energy level structure of the $\Lambda$-system, which consists of two ground states labeled as $|1\rangle$ and $|2\rangle$, and an excited state labeled as $|3\rangle$. We assume that a laser drives the transition $|1\rangle \leftrightarrow |3\rangle$ with a Rabi frequency $\Omega$ and a detuning $\Delta$, and generates Raman emission from the transition $|3\rangle \rightarrow |2\rangle$. In a rotating reference frame with respect to the driving laser, the Hamiltonian of the system is given by

$$\hat{H} = \hbar \Delta \hat{\sigma}_{33} + \frac{\hbar}{2} \Omega (\hat{\sigma}_{13} + \hat{\sigma}_{31}), \quad (S9)$$

where the operator $\hat{\sigma}_{ij}$ is defined as $\hat{\sigma}_{ij} = |i\rangle \langle j| + \text{h.c.}$ for $i, j \in \{1, 2, 3\}$. The dynamics of the system expectation values can be derived from the Heisenberg-Langevin equations, resulting in

$$\frac{d}{dt} \langle \hat{\sigma}_{13} \rangle = -(i\Delta + \gamma) \langle \hat{\sigma}_{13} \rangle - i \frac{\Omega}{2} \left( \langle \hat{\sigma}_{11} \rangle - \langle \hat{\sigma}_{33} \rangle \right), \quad (S10)$$

$$\frac{d}{dt} \langle \hat{\sigma}_{33} \rangle = -\gamma_{\text{tot}} \langle \hat{\sigma}_{33} \rangle - i \frac{\Omega}{2} \left( \langle \hat{\sigma}_{31} \rangle - \langle \hat{\sigma}_{13} \rangle \right), \quad (S11)$$
\[
\frac{d\langle \hat{\sigma}_{22} \rangle}{dt} = \Gamma \langle \hat{\sigma}_{33} \rangle. \quad \text{(S12)}
\]

In Eqs. (S10) - (S12), \( \gamma \) is the dipole decoherence rate of transition \(|1\rangle \leftrightarrow |3\rangle\), \( \Gamma \) is the spontaneous emission rate of the transition \(|3\rangle \rightarrow |2\rangle\), \( \Gamma_{\text{tot}} \) is the total decay rate of the excited state \(|3\rangle\). In the large detuning limit where \( \Omega, \Gamma_{\text{tot}}, \gamma \ll \Delta \), the excited state \(|3\rangle\) is weakly excited. Thus, we can adiabatically eliminate this state by taking the steady-state solution of Eqs. (S10) and (S11). Substituting the steady-state solutions into Eq. (S12), we obtain that

\[
\frac{d\langle \hat{\sigma}_{22} \rangle}{dt} = \Gamma \cdot \frac{\alpha}{2\alpha + \Gamma_{\text{tot}}} \left( 1 - \langle \hat{\sigma}_{22} \rangle \right), \quad \text{(S13)}
\]

where \( \alpha = \frac{\Omega^2}{2} \frac{\gamma}{\Delta^2 + \gamma^2} \). To obtain Eq. (S13), we used the identity that \( \langle \hat{\sigma}_{11} \rangle + \langle \hat{\sigma}_{22} \rangle + \langle \hat{\sigma}_{33} \rangle = 1 \) and \( \langle \hat{\sigma}_{33} \rangle = \langle \hat{\sigma}_{13} \rangle^\dagger \).

Figure S2. Schematics of Raman emission process for a bare emitter (a) and an emitter that couples to the cavity (b) respectively.

Eq. (S13) has a very clear physical interpretation. It shows that the population of the
ground state $|2\rangle$ grows exponentially with a rate $R_0 = \Gamma \cdot \frac{\alpha}{2\alpha + \Gamma_{tot}}$, which is exactly the Raman emission rate. To simplify the expression of the Raman emission rate, we assume the ideal scenario where the excited state $|3\rangle$ decays only to the ground state $|2\rangle$ through spontaneous emission ($\Gamma_{tot} = \Gamma$), and the linewidth of transition $|1\rangle \leftrightarrow |3\rangle$ is lifetime limited ($\gamma = \frac{\Gamma}{2}$). This assumption gives the upper bound of the Raman emission rate. Under this assumption, the Raman emission rate is given by $R_0 = \Gamma \cdot \frac{\alpha}{2\alpha + \Gamma}$, where $\alpha = \frac{\Omega^2 \cdot \Gamma}{4 \Delta^2 + (\Gamma/2)^2}$.

In the large detuning limit where $\Omega, \Gamma \ll \Delta$, we have $\alpha \ll \Gamma$, thus the Raman emission rate is given by $R_0 = \alpha = \frac{\Omega^2}{4\Delta^2} \Gamma$.

Now we calculate the rate of the cavity enhanced Raman emission. As explained in the main text, by adiabatic elimination of the state $|e, 0\rangle$, the cavity enhanced Raman emission can be understood as an effective Rabi oscillation between $|g_1, 0\rangle$ and $|g_2, 1\rangle$, followed by a decay from $|g_2, 1\rangle$ to $|g_2, 0\rangle$ with a rate $\kappa$. Figure S2(b) shows this simplified picture. This picture resembles the three-level systems shown in Fig. S2(a), if we define $|1\rangle \equiv |g_1, 0\rangle$, $|2\rangle \equiv |g_2, 0\rangle$ and $|3\rangle \equiv |g_2, 1\rangle$. Therefore, we could calculate the cavity enhanced Raman emission rate $R_c$ following the same derivations shown above, given by $R_c = \frac{\Omega_{eff}^2}{\kappa}$, where $\Omega_{eff}$ is given by $\Omega_{eff} = \Omega g / \Delta$ as shown in Sec. 2.
4. Details of measurement setups and techniques

4.1 Complete schematic of the measurement setup

Figure S3 shows a schematic of the optical setup we used for all our reported measurements. We mount the sample in a closed-cycle cryostat (Montana instruments) and cool it down to 4 K. We use a confocal microscope with an objective lens that has a numerical aperture of 0.9 to perform sample excitation and collection. We use three light sources for sample excitation: a supercontinuum source which is used for characterizing the transmission spectrum of the cavity (Fig. 2(b) of the main text), a pulsed Ti: Sapphire laser which is used for lifetime measurement (Fig. 2(d) of the main text), and a tunable continuous-wave Ti: Sapphire laser which is used both for generating photoluminescence (Fig. 2(c) of the main text) and Raman emission (Fig. 3 of the main text). We tune the wavelength of the continuous-wave Ti: Sapphire laser to 720 nm to generate photoluminescence, and tune it to near transition $|g\rangle \leftrightarrow |e\rangle$ to generate Raman emission. We couple the collected signal either to a single mode fiber, or directly to a spectrometer through free-space. We use a monochromator to reject the laser reflection for the Raman emission measurements.

4.2 Cavity transmission measurement

To obtain the cavity transmission spectrum as shown in Fig. 2(b) of the main text, we use a broadband supercontinuum laser to excite the notch located at one end of the nanobeam. The notch is designed to couple light from free-space into the waveguide (see Sec. 8 for detailed
information). The second notch at the other end of the nanobeam scatters the transmitted signal back into free space, which we collect with a single mode fiber. We send the fiber-coupled signal to the spectrometer to record the cavity transmission spectrum.

![Figure S3](image)

**Figure S3.** Complete schematics of the measurement setup. M, mirror; FM, flip mirror; SPCM, single-photon counting module.

### 4.3 Photoluminescence spectrum measurement

Figure 2(c) of the main text shows the measured photoluminescence spectrum of the SiV\(^-\) center. We directly excite the SiV\(^-\) center at the center of the nanobeam using a Ti: Sapphire
continuous-wave laser that is tuned to 720 nm, which is blue detuned from all the four optical transitions of the SiV<sup>−</sup> center (ranging from 735 to 738 nm). This is a typical way to generate photoluminescence from SiV<sup>−</sup> centers [13, 31, 36]. To eliminate the effect of the cavity on the emission properties, we red-detune the cavity by more than 40 cavity linewidths from all the four transitions of the SiV<sup>−</sup> center. We couple the collected signal directly to a spectrometer through free-space. We use a dielectric-coated band pass filter with a center wavelength of 740 nm and a bandwidth of 13 nm to spectrally reject the laser reflection before the spectrometer.

### 4.4 Time-resolved photoluminescence measurement

Figure 2(d) of the main text shows the time-resolved photoluminescence measurement which we use to extract the lifetime of the excited state. In this measurement, we use a 2-ps pulse with a center wavelength of 720 nm to excite the SiV<sup>−</sup> center at the center of the nanobeam, and detect the photoluminescence emission with a silicon avalanche photodiode single-photon detector. We use a monochromator to spectrally filter only the emission from the transition \(|g_2\rangle \leftrightarrow |e\rangle\) before the detector. The data shown in Fig. 2(d) of the main text is generated by a time-correlated single-photon counting system (PicoHarp 300) which records the delay time between the laser pulse and each detected photon.

### 4.5 Cavity-enhanced Raman emission measurement

In the cavity-enhanced Raman emission measurement (Fig. 3 of the main text), we excite the
device using a Ti: Sapphire continuous-wave laser that is near resonant with transition $|g_i\rangle \leftrightarrow |e\rangle$, and collect the emission from the cavity. We couple the excitation laser through the notch of the waveguide, which is designed to couple light from free-space into the waveguide (see Sec. 8 for detailed information). We collect the cavity emission from its scattering into free-space. We couple the signal into a monochromator to spectrally reject the excitation laser, and then send it to a spectrometer (which is essentially another monochromator followed by a CCD camera).

As depicted in Fig. S3, the mechanism of the monochromator is based on spatial dispersion of light upon diffraction from a grating. After spatial dispersion, a mechanical slit is used to spectrally select the signal within certain spectral range. The filter bandwidth and center wavelength can be tuned by adjusting the width and the position of the slit. In our measurement, we align the monochromator filter to be centered at the transition $|g_i\rangle \leftrightarrow |e\rangle$, with a filter bandwidth of 120 GHz. We measure a maximum transmission efficiency of the monochromator to be 17.4%. The transmission efficiency drops to 0.0014% at the frequency of the laser, which is ~544 GHz blue detuned from the center of the spectral filter. Therefore the extinction ratio of the monochromator is determined to be 12400.

5. Photoluminescence of bulk SiV$^-$ center ensembles

Figure S4 shows the photoluminescence spectrum of an ensemble of SiV$^-$ centers in the bulk diamond. Here we plot the emission spectrum as a function of detuning from the peak C.
From the frequency splitting between the emission peaks C and D, we calculate the ground state splitting to be $\delta_\pi / 2\pi = 50 \text{ GHz}$ for bulk SiV$^-$ centers. This value is consistent with the values reported in many studies [13, 30-32, 36].

![Photoluminescence spectrum of an ensemble of SiV$^-$ centers in the bulk diamond.](image)

**Figure S4.** Photoluminescence spectrum of an ensemble of SiV$^-$ centers in the bulk diamond.

6. **Second order correlation measurements**

We perform second order correlation measurements to verify that the emission peaks C and D shown in Fig. 2(c) of the main text originate from a single silicon-vacancy (SiV$^-$) center. We excite the device using a 2-ps pulsed laser with a repetition rate of 80 MHz and a center wavelength of 720 nm. We collect the emission using a multi-mode fiber and send it to a Hanbury Brown-Twiss (HBT) intensity interferometer composed of a 50/50 beam-splitter and two Single Photon Counting Modules (SPCMs). We use a time correlated single-photon counting system (PicoHarp 300) to process the detection events from the two SPCMs and obtain the second order correlation. To isolate the emission from peaks C and D respectively, we resonantly couple the cavity with either peak C or D and use the cavity as a spectral filter. Figure S5 shows
the measured second order correlations when the cavity is resonant with peak C and D respectively, where $\tau$ is the delay time between two detection events obtained by the two SPCMs. We observe strong suppression of the second order correlation near $\tau = 0$ in both cases, confirming that both emission peaks are from a single SiV$^-$ center.

Figure S5. Second order correlations of the SiV$^-$ emission when the cavity is resonant with emission peak C and D respectively.

7. Estimation of Purcell factor

We estimate the Purcell factor $F$ based on the calculated coupling strength, defined as

$$F = \frac{4g^2/\kappa}{\Gamma_{\text{bare}}}$$

where $\Gamma_{\text{bare}}$ is the spontaneous emission rate of transition $|e\rangle \rightarrow |g_2\rangle$ when it decouples with the cavity. It is not straightforward to obtain the value of $\Gamma_{\text{bare}}$, since we can only measure the lifetime (or total decay rate) of the excited state $|e\rangle$ when the cavity is far detuned, which includes decay through non-radiative processes, phonon-sideband emissions, and spontaneous emission into the other zero-phonon line $|e\rangle \rightarrow |g_1\rangle$. To obtain a lower bound of the
Purcell factor, here we estimate a higher bound of $\Gamma_{\text{bare}}$. We assume a higher-bound quantum yield of $\eta_{\text{radiative}} = 30\%$ and a zero-phonon-line emission fraction of $\eta_{\text{ZPL}} = 80\%$. These two values together have accounted for the fraction of decay through emissions into the zero-phonon-lines. However, the excited state $|e\rangle$ could still decay through two possible ZPLs, $|e\rangle \rightarrow |g_1\rangle$ and $|e\rangle \rightarrow |g_2\rangle$. We estimate the fraction of the zero-phonon-line emission into transition $|e\rangle \rightarrow |g_2\rangle$ to be $\eta_D = 10\%$ based on the measured photoluminescence spectrum shown in Fig. 2(c) of the main text. Therefore, we could calculate the higher-bound of $\Gamma_{\text{bare}}$ given by $\Gamma_{\text{bare}} = \eta_{\text{radiative}} \cdot \eta_{\text{ZPL}} \cdot \eta_D \frac{1}{\tau_{\text{off}}} = 2\pi \times 2.1 \text{ MHz}$. We thus estimate the lower-bound of the Purcell factor to be $F = \frac{4g^2/\kappa}{\Gamma_{\text{bare}}} = 22.7$.

**Figure S6.** Photoluminescence spectrum of the emitter when the cavity is resonant (red) and far detuned (blue) from the transition $|e\rangle \rightarrow |g_2\rangle$.

To further confirm the lower-bound value of the Purcell factor, we measure the photoluminescence of the emitter when the cavity is on resonance and far detuned from
transition $|e\rangle \rightarrow |g_2\rangle$, shown as the red and blue solid lines in Fig. S6. The emission intensity from transition $|e\rangle \rightarrow |g_2\rangle$ (peak D) increases by a factor of 60 when the cavity is resonant, which is another supportive evidence of a strong Purcell enhancement.

8. **Design of the notches for free-space to waveguide coupling**

In our experiment, to reject the direct reflection of the laser from the sample surface, we spatially separate the excitation and collection by irradiating the laser at a notch located at the end of the nanobeam, which is designed for free-space to waveguide coupling. Figure S7(a) shows the dimensions of the notch, which follows a similar design as shown in Ref. [13]. Figure S7(b) shows a scanning electron microscope image of the notch after fabrication.

![Figure S7](image)

**Figure S7.** (a) Dimensions of the notch that is designed for coupling light from free-space into the nanobeam waveguide. (b) Scanning electron microscope image of the notch after fabrication.

9. **Complete derivation of system master equation**

We first derive the Lindblad operator for the electron-phonon interaction Hamiltonian $\hat{H}_{\text{sys–bath}}$. To do that we have to first write the interaction Hamiltonian $\hat{H}_{\text{sys–bath}}$ in the diagonal
basis of $\hat{H}_{\text{sys}}$ (Eq. (1) of the main text), and then write it in the rotating reference frame with respect to $\hat{H}_{\text{sys}} + \hat{H}_{\text{bath}}$. In the diagonal basis, we can write $\hat{H}_{\text{sys}}$ as

$$\hat{H}_{\text{sys}} = \omega_+ \langle + | + \omega_- \langle - | + \omega_d | d \rangle \langle d |,$$

(S14)

where the eigenstates $|+\rangle$, $|−\rangle$, and $|d\rangle$ are given by Eqs. (5) – (7) in the main text, the eigenfrequencies $\omega_+$, $\omega_−$, and $\omega_d$ are given by $\omega_+ = \Delta + \frac{(\Omega/2)^2 + g^2}{\Delta}$, $\omega_− = -\frac{(\Omega/2)^2 + g^2}{\Delta}$, and $\omega_d = 0$ respectively. In the rotating reference frame with respect to $\hat{H}_{\text{sys}} + \hat{H}_{\text{bath}}$, we could rewrite $\hat{H}_{\text{sys-bath}}$ as

$$\hat{H}_{\text{sys-bath}} = \sqrt{\frac{g^2 + (\Omega/2)^2}{\Delta}} \sum_k b_k \left( x_k |+\rangle \langle +| e^{i(\Lambda_1 - \omega_k)\tau} + y_k |+\rangle \langle d| e^{i(\Lambda_2 - \omega_d)\tau} \right) + \text{h.c.},$$

(S15)

where $\Lambda_1$ and $\Lambda_2$ are given by $\Lambda_1 = \omega_+ - \omega_−$ and $\Lambda_2 = \omega_+ - \omega_d$ respectively, $x_k$ and $y_k$ are given by $x_k = p_k \frac{(\Omega/2)^2 + q_k}{g^2 + (\Omega/2)^2} - r_k$ and $y_k = (p_k - q_k) \frac{g(\Omega/2)}{g^2 + (\Omega/2)^2}$ respectively. To obtain Eq. (S15), we have utilized the rotating wave approximation to keep only the slowly varying terms. We eliminate the phonon coupling terms with the operators $|+\rangle \langle +|$, $|−\rangle \langle −|$, and $|d\rangle \langle d|$ because they interact with phonons at zero frequency where phonon density of states vanishes. Similarly, we eliminate the phonon coupling terms with the operators $|d\rangle \langle −|$, and $|−\rangle \langle d|$ because they interact with phonons at a low frequency near $\omega_d = \omega_+ + \frac{(\Omega/2)^2 + g^2}{\Delta}$, which is in the order of 100 MHz. Such frequencies correspond to a phonon wavelength longer than 10 µm, which cannot exist in our nanobeam structure.

Now we can derive a master equation by integrating the von Neumann equation for the
density matrix $\rho$ of the joint system and phonon bath, and then tracing over the phonon modes, given by

$$
\frac{d\rho_{\text{sys}}}{dt} = -\int_0^t \text{tr}_{\text{bath}} \left( \left[ \hat{H}_{\text{sys-bath}}(t), \left[ \hat{H}_{\text{sys-bath}}(t'), \rho(t') \right] \right] \right) dt'.
$$

(S16)

We make the Born-Markov approximation, which allows us to substitute $\rho(t')$ with $\rho(t)$ and write it as $\rho = \rho_{\text{sys}} \otimes \rho_{\text{bath}}$. These assumptions result in a master equation given by

$$
\frac{d\rho_{\text{sys}}}{dt} = -\int_0^t \text{tr}_{\text{bath}} \left( \left[ \hat{H}_{\text{sys-bath}}(t), \left[ \hat{H}_{\text{sys-bath}}(t), \rho_{\text{sys}}(t) \otimes \rho_{\text{bath}} \right] \right] \right) dt'.
$$

(S17)

We further rewrite $\hat{H}_{\text{sys-bath}}$ as $\hat{H}_{\text{sys-bath}} = \hat{H}_{\text{sys-bath}}^{(1)} + \hat{H}_{\text{sys-bath}}^{(2)}$, where $\hat{H}_{\text{sys-bath}}^{(1)}$ and $\hat{H}_{\text{sys-bath}}^{(2)}$ are given by

$$
\hat{H}_{\text{sys-bath}}^{(1)} = \frac{\sqrt{g^2 + (\Omega/2)^2}}{\Delta} \sum_k \chi_k b_k |+\rangle \langle -| e^{i(\Lambda_1 - \epsilon_k)} + \text{h.c.},
$$

(S18)

$$
\hat{H}_{\text{sys-bath}}^{(2)} = \frac{\sqrt{g^2 + (\Omega/2)^2}}{\Delta} \sum_k \gamma_k b_k |+\rangle \langle d| e^{i(\Lambda_2 - \epsilon_k)} + \text{h.c.}.
$$

(S19)

Since $\hat{H}_{\text{sys-bath}}^{(1)}$ and $\hat{H}_{\text{sys-bath}}^{(2)}$ involve interaction with phonons of different frequencies separated by $\sim 100$ MHz, they cannot interact with the same phonon mode. Therefore, we could further rewrite Eq. (S17) as

$$
\frac{d\rho_{\text{sys}}}{dt} = -\sum_{m=1}^2 \int_0^t \text{tr}_{\text{bath}} \left( \left[ \hat{H}_{\text{sys-bath}}^{(m)} \left[ \hat{H}_{\text{sys-bath}}^{(m)}(t), \rho_{\text{sys}}(t) \otimes \rho_{\text{bath}} \right] \right] \right) dt'.
$$

(S20)

This leads to the final master equation given by $\frac{d\rho_{\text{sys}}}{dt} = -i \left[ \hat{H}_{\text{sys}}, \rho_{\text{sys}} \right] + \sum_{m=1}^2 L_{\text{phonon}}^{(m)}\left( \rho_{\text{sys}} \right)$, where $L_{\text{phonon}}^{(1)}\left( \rho_{\text{sys}} \right)$ and $L_{\text{phonon}}^{(2)}\left( \rho_{\text{sys}} \right)$ are given by
\begin{equation}
L_{phonon}^{(1,2)}(\rho_{sys}) = \frac{g^2 + (\Omega/2)^2}{\Delta^2} J_{1,2}(\Lambda_{1,2}) \left[ n_{th}(\Lambda_{1,2}) D(|+\rangle\langle-|) + (1 + n_{th}(\Lambda_{1,2})) D(|-\rangle\langle+|) \right],
\end{equation}

where $J_{1,2}(\Lambda_{1,2})$ is the phonon spectral density given by $J_i(\Lambda_i) = 2\pi \sum_k |x_k|^2 \delta(\omega_k - \Lambda_i)$ and $J_2(\Lambda_2) = 2\pi \sum_k |y_k|^2 \delta(\omega_k - \Lambda_2)$ respectively, and $n_{th}(\Lambda_{1,2})$ is the number of phonons per mode, which follows the Bose-Einstein distribution given by $n_{th}(\Lambda_{1,2}) = \left[ \exp(\Lambda_{1,2}/k_B T) - 1 \right]^{-1}$.

Since both the phonon spectral density $J_{1,2}(\Lambda_{1,2})$ and the thermal distribution function $n_{th}(\Lambda_{1,2})$ are relatively flat as a function of phonon frequency, and $\Lambda_{1,2} \approx \Delta$ in the limit $\Omega, g << \Delta$, we could approximately write $J_{1,2}(\Lambda_{1,2})$ and $n_{th}(\Lambda_{1,2})$ as $J_{1,2}(\Lambda_{1,2}) = J_{1,2}(\Delta)$ and $n_{th}(\Lambda_{1,2}) = n_{th}(\Delta)$ respectively. This gives the phonon dissipator provided as Eq. (4) in the main text.

In our numerical simulation, we use the full master equation that account for all possible dissipation mechanisms, given by

$$
\frac{d\rho_{sys}}{dt} = -i[\hat{H}_{sys}, \rho_{sys}] + L_{phonon}(\rho_{sys}) + L_{cav}(\rho_{sys}) + L_{SiV}(\rho_{sys}),
$$

where $L_{cav}(\rho_{sys}) = \kappa \hat{D}(\hat{a})$ is the cavity decay, and $L_{SiV}(\rho_{sys})$ is the decay of the SiV\(^-\) center, given by

$$
L_{SiV}(\rho_{sys}) = \gamma_1 \hat{D}(\ket{g_1}\bra{e}) + \gamma_2 \hat{D}(\ket{g_2}\bra{e}) + \gamma_{flip} \hat{D}(\ket{g_1}\bra{g_2}).
$$

In Eq. (S22), $\gamma_1$ and $\gamma_2$ are decay rates from the excited state $|e\rangle$ to the ground states $|g_1\rangle$ and $|g_2\rangle$ respectively, $\gamma_{flip}$ is the decay rate from $|g_2\rangle$ to $|g_1\rangle$. We do not include the state flipping from $|g_1\rangle$ to $|g_2\rangle$ in the Liouvillian superoperator, because this process requires absorption of a phonon at $\Delta / 2\pi = 544 \text{ GHz}$ that is much larger than the value $k_B T / 2\pi = 83 \text{ GHz}$, and therefore is much slower than its reverse process.
We numerically solve the master equation of the system and calculate the cavity emission spectrum using the quantum regression theorem. We fix $g$ and $\kappa$ using experimentally measured values given by $g/2\pi = 0.80\,\text{GHz}$ and $\kappa/2\pi = 53.7\,\text{GHz}$. We assume that the decay rates from the excited state $|e\rangle$ to the ground states $|g_1\rangle$ and $|g_2\rangle$ are identical, thus we have $\gamma_1 = \gamma_2 = \frac{1}{\tau_{\text{off}}} = 2\pi \times 0.046\,\text{GHz}$. This assumption is valid since the dominant decay mechanism of the excited state is through a non-radiative process [37], which rate is irrelevant of the final ground state. To determine the driving Rabi frequency $\Omega$, we resonantly drive transition $|g_1\rangle \leftrightarrow |e\rangle$, and measure the fluorescence intensity as we vary the driving laser power. This measurement allows us to obtain the saturation power for transition $|g_1\rangle \leftrightarrow |e\rangle$, enabling us to determine the driving Rabi frequency based on the measured laser power. In our experiment, we use a driving Rabi frequency of $\Omega/2\pi = 2.58\,\text{GHz}$. The only parameter we cannot determine is $\gamma_{\text{flip}}$. However, as we verified numerically, the value of $\gamma_{\text{flip}}$ only determines the number of excitation and emission cycles per second – it does not affect the R/S ratio. In the calculation we simply fix $\gamma_{\text{flip}}$ to be $\gamma_{\text{flip}}/2\pi = 0.8\,\text{GHz}$ based on an estimate from a previous literature [31].