Double Cavity Exceptional Point Enhanced Laser Cooling of Semiconductors

Haonan Chang (常浩男),[1,2] Feilong Song (宋飞龙),[1] and Jun Zhang (张俊)[1,2,*]

[1] *State Key Laboratory of Semiconductor Physics and Chip Technologies, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China*

[2] *Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China*

Laser cooling of semiconductors requires a high external quantum efficiency for phonon-assisted upconversion while suppressing heat generated by background absorption and non-radiative relaxation. Nanophotonic resonances can strengthen the upconversion channel, but in a single-mode structure may also enhance parasitic pump heating. We introduce a two-cavity non-Hermitian architecture that separates preparation of a lower reservoir state from extraction of the phonon-promoted radiative state. A passive exceptional point, containing no optical gain, is placed in the output branch, where it reshapes the projected final-state optical response; all anti-Hermitian diagonal terms describe radiative leakage, internal absorption, or material decay. For a CdS parameter set, the four-mode calculation gives a 1.23×10^3 enhancement of upconverted photoluminescence (UCPL) relative to a single-mode control, while the pump-induced heat load rises by only 2.8×10^{-3}%. The gain consists of a factor of 3.45×10^2 from the independent output resonance and a further factor of 3.56 from exceptional-point tuning.

Keywords: laser cooling, semiconductors, phonon-assisted upconversion photoluminescence, double cavity, exceptional point

Optical refrigeration removes lattice energy through anti-Stokes fluorescence: lower-frequency pump photons are absorbed, the resulting excitations acquire thermal energy from the lattice, and higher-frequency fluorescence photons subsequently escape from the material [1]. Optical refrigeration is therefore also attractive for cryogenic-chip cooling as a compact, vibration-free approach that complements conventional refrigeration technologies [2–4]. The energy removed per cooling cycle is the mean photon-energy increase $\Delta E = \hbar(\bar{\omega}_f - \omega_p)$, which represents an ensemble-averaged transfer of thermal energy. Rare-earth-doped solids have been optically cooled into the cryogenic regime, demonstrating that net refrigeration is possible when the anti-Stokes cooling power exceeds the combined heat generated by nonradiative relaxation and parasitic absorption [5–11]. The same heat-balance requirements apply to both rare-earth-doped crystals and semiconductors: a sufficiently large anti-Stokes energy shift increases the heat removed per escaping photon and relaxes the required external quantum efficiency, while high radiative efficiency and weak background absorption suppress competing heat generation [12, 13]. Compared with rare-earth ions, semiconductors provide much stronger band-edge and excitonic absorption, large optical oscillator strengths, and compatibility with nanophotonic structures that enable strong light–matter interaction and integrated optical control [14–19]. They also introduce additional challenges, including nonradiative carrier recombination, carrier–phonon thermalization, and inefficient photon escape caused by their high refractive indices. Laser cooling of semiconductors therefore requires a high external quantum efficiency, sufficiently large mean anti-Stokes energy extraction, and extremely weak parasitic absorption.

Nanophotonic resonances provide a route to improve the optical part of the refrigeration cycle by modifying the photonic density of states, enhancing radiative emission, and directing fluorescence into efficiently escaping optical channels [20–22]. Related photonic-density-of-states engineering has also been proposed for Raman cooling of solids [23]. An ordinary single-resonance architecture, however, provides only one optical mode for both reservoir preparation and anti-Stokes extraction. If the same resonance is required to prepare the low-energy reservoir at ω_p and to extract the anti-Stokes emission at ω_f, its resonance frequency, linewidth, field overlap, and external coupling cannot be optimized independently for the two processes. A high-Q resonance with strong material overlap can increase the intracavity pump field and reservoir-preparation rate, but the longer photon lifetime also increases the cumulative probability of parasitic absorption. Conversely, strong radiative overcoupling facilitates photon escape but generally reduces intracavity field buildup. The resulting external upconversion efficiency, which is determined by reservoir preparation, thermally assisted promotion, radiative conversion, and photon escape, is therefore restricted by a trade-off between the initial excitation and final extraction stages. Exceptional points (EPs) provide a form of optical control beyond that available from an ordinary resonance and have been exploited in nanophotonic systems for mode-selective lasing, asymmetric mode conversion, enhanced spectral responses, and the control of spontaneous emission and radiative decay [24–33]. At an EP, the complex eigenfrequencies and eigenvectors of two coupled lossy modes coalesce. In a passive nanophotonic system, this modal coalescence can produce a non-

Lorentzian spectral response and a finite reshaping or enhancement of the local photonic density of states and spontaneous-emission rate, provided that the material excitation couples to the coalesced modes and the selected radiation channel. An EP can modify the final radiative response so that a thermally promoted excitation is more likely to be converted into an escaping anti-Stokes photon rather than lost through nonradiative or optically trapped channels.

Here we separate the pump-preparation and photon-extraction functions using a dual-cavity architecture and introduce the non-Hermitian enhancement only into the output branch. The architecture follows directly from the phonon-assisted cooling cycle in CdS, as illustrated in Fig. 1(a): a 514-nm pump populates the lower reservoir manifold L, absorption of one approximately 37-meV CdS LO phonon promotes the excitation to the upper radiative manifold U, and the excitation subsequently emits a higher-energy photon near 506 nm. To separate preparation from extraction, two distinct optical modes are introduced, as shown in Fig. 1(b). The pump mode C_p couples predominantly to L, whereas the spectrally distinct output mode C_f couples predominantly to U, corresponding to $g_{pL} \gg g_{pU}$ and $g_{fU} \gg g_{fL}$. Thus the reduced model contains two optical modes, C_p and C_f, and two material excitation manifolds, L and U. The functional separation of the two branches is emphasized in Fig. 1(c): the C_p branch prepares the lower reservoir, whereas the C_f branch extracts the phonon-promoted excitation as an anti-Stokes photon. The passive exceptional point is formed only in the coupled C_f–U output subsystem, so that the EP-associated resolvent modification acts predominantly on the final-state extraction process without comparably modifying the pump response. A corresponding physical implementation is shown in Fig. 1(d), where spatially distinct CdS and SiO_2 microdisks provide the pump and output resonators, respectively, and their relative geometry controls the overlap between the output mode and the CdS active region. The pump and output modes can therefore be optimized independently for reservoir preparation and radiative extraction. The system is entirely passive rather than gain–loss balanced, and no positive imaginary gain term is introduced. For the CdS parameter set considered below, the four-mode model predicts a system-level UCPL enhancement of 1.23×10^3 relative to a single-mode control, while the modeled pump-induced heat increases by only $2.8 \times 10^{-3}\%$.

Consistent with the four-mode representation in Fig. 1(b), we use the ordered modal-amplitude basis (C_p, C_f, U, L). In the reduced Hamiltonian and Green function, the unsubscripted frequency argument ω is measured relative to the pump, so that the pump corresponds to $\omega = 0$ and the nominal one-phonon anti-Stokes output corresponds to $\omega = \Omega_\nu$. By contrast, ω_p and ω_f denote the absolute pump and emitted optical frequencies, respectively, as used in the photon-

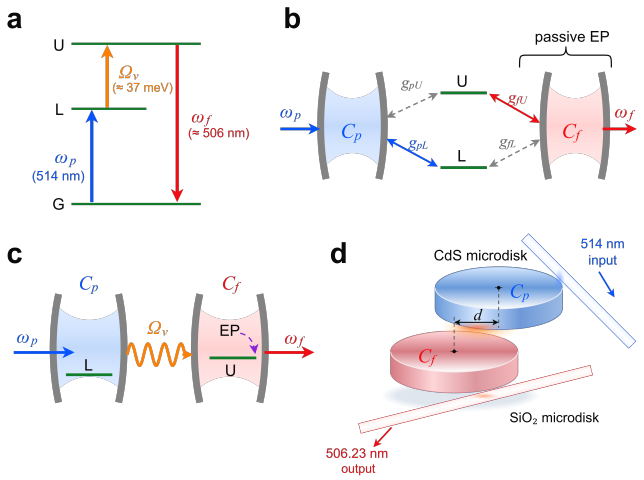

FIG. 1. Operating principle and physical implementation of the dual-cavity CdS cooling architecture. (a) One-phonon anti-Stokes cooling cycle. A 514-nm pump prepares the lower reservoir manifold L, absorption of an approximately 37-meV CdS LO phonon promotes the excitation to the upper radiative manifold U, and the excitation subsequently emits an anti-Stokes photon near 506 nm. (b) Reduced four-mode coupling scheme. The pump mode C_p couples predominantly to L, with $g_{pL} \gg g_{pU}$, whereas the output mode C_f couples predominantly to U, with $g_{fU} \gg g_{fL}$; the passive exceptional point is formed in the coupled C_f–U output subsystem. (c) Dual-cavity operating principle, in which the pump and anti-Stokes extraction processes are assigned to spatially and spectrally distinct optical modes. The pump branch prepares L, phonon absorption transfers the excitation to U, and the output branch extracts the resulting anti-Stokes photon. The material manifolds L and U remain distinct from the optical cavity modes C_p and C_f. (d) Schematic physical realization using vertically stacked and laterally offset CdS and SiO_2 microdisks. Independent input and output waveguides are evanescently coupled to the corresponding pump and output microdisks, while the lateral offset d controls the modal overlap between the output mode and the CdS active region.

energy and cooling-efficiency expressions. Here C_p and C_f label the pump and output optical modes, while U and L label effective upper-radiative and lower-reservoir material excitation manifolds. Thus, C_p and C_f denote photonic cavity modes, whereas L and U denote material excitation manifolds; these four components should not be interpreted as four equivalent energy levels. Equation (1) is therefore an effective non-Hermitian coupled-mode dynamical matrix rather than a many-body quantum-state Hamiltonian. With the amplitude vector $\boldsymbol{x} = (x_{C_p}, x_{C_f}, x_U, x_L)^T$, the optical and material dynamics are described by

$$H_0 = \begin{pmatrix} -i\kappa_p/2 & 0 & g_{pU} & g_{pL} \\ 0 & \Omega_\nu + \Delta_f - i\kappa_f/2 & g_{fU} & g_{fL} \\ g_{pU} & g_{fU} & \Omega_\nu - i\gamma_U/2 & 0 \\ g_{pL} & g_{fL} & 0 & -i\gamma_L/2 \end{pmatrix}. \tag{1}$$

The Hamiltonian in Eq. (1) defines a purely passive non-Hermitian system. It contains no optical-gain contri-

bution of the form $+iG/2$; all diagonal anti-Hermitian terms have negative imaginary parts, $-i\kappa_j/2$ or $-i\gamma_\mu/2$. The cavity linewidths include leakage into external radiation channels and internal absorption, while the material linewidths include population decay and homogeneous broadening or dephasing, including nonradiative recombination. These open-system channels make $H_0 \neq H_0^\dagger$. External leakage through the selected output port is useful for photon extraction, whereas internal absorption and nonradiative decay return energy to the device as heat. The passive EP considered below results from coherent C_f–U coupling together with their unequal decay rates; neither an amplifying medium nor balanced gain and loss is required.

The optical linewidths are separated as $\kappa_j = \kappa_{je} + \kappa_{ji}$, where κ_{je} describes coupling to useful external ports and κ_{ji} represents internal absorption. The coupling hierarchy

$$g_{pL} \gg g_{pU}, \qquad g_{fU} \gg g_{fL} \tag{2}$$

assigns a distinct physical role to each optical mode. The pump mode loads the reservoir, whereas the output mode extracts the radiative state. The hierarchy does not require exactly vanishing cross couplings. The coupling strengths to L and U are jointly determined by spectral detuning, optical mode profiles, the CdS spectral response, and polarization selection rules. Residual pump leakage into the output channel is represented separately by ϵ. For the nominal reduced-model parameters, $g_{pU}/g_{pL} = 0.174/8.12 \simeq 2.1\%$, which quantifies the required pump-branch selectivity. Energy leaving through κ_{fe} contributes to the anti-Stokes cooling channel; energy lost through κ_{fi} is returned to the structure as heat. A broad output resonance can therefore be useful if its linewidth is predominantly external. The phonon-assisted transition is introduced separately from the non-Hermitian optical Hamiltonian,

$$V_\nu^{(+)} = \Lambda_\nu \sqrt{n_\nu(T)}\, |U\rangle\langle L|, \qquad n_\nu = (e^{\hbar\Omega_\nu/k_BT} - 1)^{-1}. \tag{3}$$

This form makes the physical sequence explicit: the excitation occupies L, annihilates one lattice phonon, and is promoted to U. For pump injection and collection through the output port, the anti-Stokes amplitude is

$$A_{\mathrm{UC}} = \boldsymbol{K}_f^\dagger G_0(\Omega_\nu) V_\nu^{(+)} G_0(0) \boldsymbol{K}_p, \tag{4}$$

where $G_0(\omega) = (\omega I - H_0)^{-1}$, $\boldsymbol{K}_p = (\sqrt{\kappa_{pe}}, \epsilon\sqrt{\kappa_{fe}}, 0, 0)^T$, and $\boldsymbol{K}_f = (0, \sqrt{\kappa_{fe}}, 0, 0)^T$. The parameter ϵ quantifies residual pump leakage into the output mode. Equation (4) factorizes the full conversion into three steps: preparation by $G_0(0)\boldsymbol{K}_p$, phonon absorption through $V_\nu^{(+)}$, and radiative extraction by $\boldsymbol{K}_f^\dagger G_0(\Omega_\nu)$. The phonon matrix element and Bose occupation multiply every optical configuration equally and therefore cancel from the enhancement ratios considered below. The factorization in Eq. (4) also separates the material and photonic parts of the cooling cycle. The vertex $\Lambda_\nu\sqrt{n_\nu}$ is fixed by the exciton–phonon interaction and the lattice occupation, while the two Green-function factors are controlled by optical design. In a single-mode system these factors inherit the same cavity denominator, so a resonance that increases the final-state amplitude generally also modifies the pump field and its associated absorption. In the two-color system, $G_0(0)$ and $G_0(\Omega_\nu)$ are dominated by different optical poles. The output EP can therefore alter the final-state density of optical states and the port residue without producing an equivalent change in the pump-state amplitude. This is the microscopic origin of the selectivity that is absent from a one-resonance description.

The pump-induced heat load follows from the steady-state pump amplitudes $\boldsymbol{x}_p = G_0(0)\boldsymbol{K}_p$,

$$\begin{aligned}\frac{P_{\mathrm{heat}}}{\hbar\omega_p |s_{\mathrm{in}}|^2} &= \kappa_{pi}|x_{C_p}|^2 + \kappa_{fi}|x_{C_f}|^2 \\ &\quad + \gamma_U^{\mathrm{nr}}|x_U|^2 + \gamma_L^{\mathrm{nr}}|x_L|^2 + \alpha_b.\end{aligned} \tag{5}$$

The added output branch enters the pump heat mainly through $\kappa_{fi}|x_{C_f}|^2$. Spectral separation, polarization selectivity, and a small ϵ suppress this term. The same output mode can nevertheless couple strongly at ω_f, where it is intentionally resonant with U. The two-cavity model thus removes the algebraic constraint that ties pump absorption and anti-Stokes extraction in a single-mode description. Each degree of freedom in Eq. (1) has a distinct role. The two states are required because cooling proceeds through a phonon-assisted transition between a lower reservoir and a higher radiative state. The two optical modes are required because preparation and extraction impose different conditions on field orientation, linewidth, and internal loss. Eliminating C_f restores the single-mode correlation between pump absorption and emission. Eliminating C_p forces the output resonance to absorb the pump directly. If the corresponding non-Hermitian degeneracy were instead introduced into the pump-related subsystem, the pump-sector susceptibility $\langle L|G_0(0)\boldsymbol{K}_p\rangle$ could be enhanced near the pump frequency, thereby increasing reservoir loading; the same enhancement would also increase the intracavity pump field and the associated internal absorption. The four-mode architecture is therefore the smallest passive model that allows the EP to act on photon escape while leaving pump loading approximately unchanged. The reservoir state also has a specific physical function. It stores the pump-created population until the inelastic transition occurs. A narrow reservoir linewidth improves spectral discrimination between direct pump absorption and the phonon-assisted pathway, whereas a broad reservoir reduces that distinction. The cooling cycle is therefore resolved into storage, phonon absorption, and radiative es-

cape rather than treated as a single resonantly enhanced scattering event.

If the spectator couplings g_{pU} and g_{fL} are neglected, the (C_f, U) block reaches a passive second-order EP [26, 28] at

$$\Delta_f = 0, \qquad g_{fU}^{\mathrm{EP}} = \frac{|\kappa_f - \gamma_U|}{4}. \tag{6}$$

The weak pump-sector couplings shift the exact location slightly, so we determine the EP from the discriminant of the full four-mode Hamiltonian. At a second-order EP the resolvent of the output block has the local Jordan expansion [24, 26, 30]

$$G_{\mathrm{out}}(\omega) \simeq \frac{\mathcal{P}_2}{(\omega - \lambda_{\mathrm{EP}})^2} + \frac{\mathcal{P}_1}{\omega - \lambda_{\mathrm{EP}}} + G_{\mathrm{reg}}(\omega), \tag{7}$$

where $\mathcal{P}_2$ connects the eigenvector and associated vector of the defective subspace. The anti-Stokes extraction factor $\boldsymbol{K}_f^\dagger G_0(\Omega_\nu)|U\rangle$ samples this enhanced pole structure. The preparation factor $\langle L|G_0(0)\boldsymbol{K}_p\rangle$ remains regular because the pump is carried by a different optical mode at a different frequency. Because the pump and output channels are separated in frequency and port projection, the EP-induced modification of the resolvent is projected predominantly onto the final-state extraction amplitude rather than the pump-preparation amplitude. A conventional final-state cavity produces a simple pole in the extraction amplitude. The EP adds a second-order contribution and simultaneously changes the radiative residue seen by the output port. Neither effect is useful if the coalesced mode is dominated by internal loss or if the pump directly excites the same subspace. The enhancement therefore depends on both the non-Hermitian pole and its port projections. In scattering language, ordinary cavity enhancement increases the probability that the system occupies an optical mode; selective extraction increases the conditional probability that a phonon-promoted excitation leaves through the desired port. Only the latter directly removes lattice energy from the device.

The second-order term in Eq. (7) should not be interpreted as a divergence of a physical passive system. Finite linewidths, imperfect tuning, saturation, and dephasing keep the response finite. Its significance is that the spectral weight carried by two simple poles is reorganized into a defective subspace whose response varies more rapidly with detuning and whose residue can project strongly onto the radiative port. The enhancement is consequently controlled by three ingredients: proximity to the EP, overlap of U with the defective eigenvector, and overlap of that eigenvector with the external channel $\boldsymbol{K}_f$. A coalescence with poor port overlap, or one dominated by internal absorption, would yield little cooling benefit even if its eigenvalues displayed a mathematically exact EP.

The eigenvalues in Figs. 2(a) and 2(b) display the expected coalescence and linewidth exchange. The simultaneous appearance of these two signatures is important: a large spectral peak without the associated non-Hermitian eigenvalue structure would describe ordinary resonance enhancement rather than the EP mechanism considered here. Fig. 2(c) shows the corresponding anti-Stokes response, whereas Fig. 2(d) provides the crucial control that the pump heat does not follow the same resonance. We choose representative passive parameters

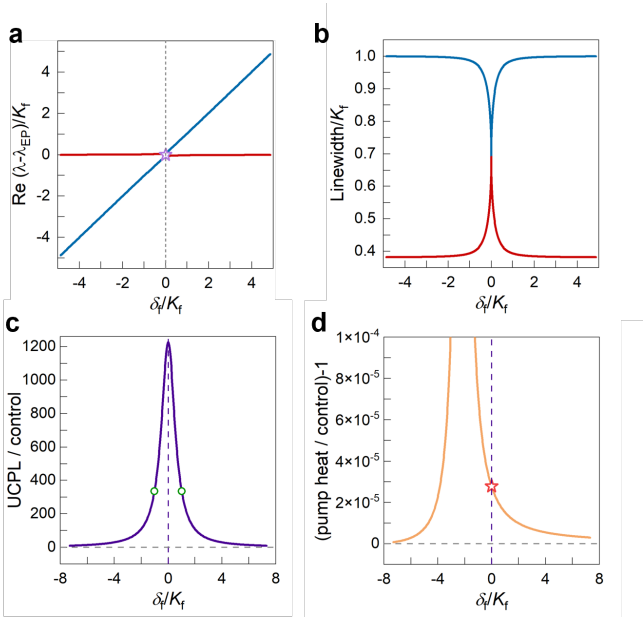

FIG. 2. Selective output response. (a) Real parts and (b) linewidths of the coalescing eigenvalues as $\delta_f = \Delta_f - \Delta_f^{\mathrm{EP}}$ crosses zero. (c) UCPL relative to the single-mode control. Open circles mark off-EP dual-mode controls at $\delta_f = \pm\kappa_f$; the star marks the EP. (d) Pump-induced heat remains nearly unchanged while the output branch is tuned through the EP.

$$(\Omega_\nu, \kappa_p, \kappa_f, \gamma_U, \gamma_L) = (37.00, 60.05, 16.45, 6.27, 3.18)\ \mathrm{meV}, \tag{8}$$

external coupling fractions $\kappa_{pe}/\kappa_p = 0.95$ and $\kappa_{fe}/\kappa_f = 0.981$, couplings $g_{pL} = 8.12$ meV, $g_{pU} = 0.174$ meV, $g_{fL} = 0$, and pump leakage $\epsilon = 0.0227$. The full-system EP occurs at $\Delta_f^{\mathrm{EP}} = 5.85 \times 10^{-4}$ meV and $g_{fU}^{\mathrm{EP}} = 2.5448$ meV. The single-mode reference retains the entire pump sector but removes C_f, so that all changes arise from adding and tuning the output branch. A separate output resonance placed one linewidth away from the EP already raises the UCPL by a factor of 3.45×10^2. This is the dual-resonance contribution: the promoted excitation is provided with a final-state optical resonance that is absent in the single-mode control. Tuning the same branch to the EP supplies an additional factor of 3.56. The total enhancement is therefore

$$\mathcal{E}_{\mathrm{UC}} = \frac{R_{\mathrm{UC}}^{\mathrm{EP}}}{R_{\mathrm{UC}}^{\mathrm{single}}} = 344.99 \times 3.56 = 1.23 \times 10^3. \tag{9}$$

The decomposition prevents the three-order-of-

magnitude system-level result from being attributed to the EP alone. The architecture first creates an independent extraction channel; the EP then reorganizes the poles and residues of that channel. The large numerical ratio in Eq. (9) is a system-level contrast rather than a thousandfold increase generated by the EP alone. The optimized hierarchy strongly suppresses the direct single-mode pathway through the upper radiative state, so the reference device has poor extraction by construction. Adding C_f restores a resonant radiative channel and accounts for most of the gain; the EP then multiplies that already selective pathway by a further factor of 3.56. This distinction is essential for physical interpretation and for experiment. The predicted advantage should be verified by comparing structures with identical pump coupling and material parameters, first without the output resonance, then with the output resonance away from the EP, and finally at the EP. Such a sequence separates architectural gain from non-Hermitian gain and prevents an artificially weak baseline from being mistaken for an EP singularity. For the same parameters,

$$\frac{P_{\mathrm{heat}}^{\mathrm{EP}}}{P_{\mathrm{heat}}^{\mathrm{single}}} = 1.000028. \tag{10}$$

A useful optical selectivity measure is consequently

$$\mathcal{S}_{\mathrm{opt}} = \frac{R_{\mathrm{UC}}^{\mathrm{EP}}/R_{\mathrm{UC}}^{\mathrm{single}}}{P_{\mathrm{heat}}^{\mathrm{EP}}/P_{\mathrm{heat}}^{\mathrm{single}}} = 1.23 \times 10^{3}. \tag{11}$$

This quantity is not a net cooling efficiency; it isolates the improvement of the useful optical channel relative to the modeled pump-induced loss. A complete refrigeration experiment must still satisfy $\hbar\Omega_\nu R_{\mathrm{UC}} > P_{\mathrm{parasitic}}$ after accounting for all absorption and non-radiative processes. Equation (11) shows that the two-cavity EP architecture changes that balance in the desired direction rather than merely increasing the overall optical intensity. The selective enhancement survives over a broad parameter range. Figure 3(a) varies the output linewidth and the undesired pump coupling to U. Large gain now requires a much sharper division of optical roles, quantified by a small direct pump coupling to U together with a narrow but still radiatively dominated output mode. The resulting high-gain region is extended rather than pointlike, so the three-order-of-magnitude enhancement is not tied to an isolated parameter spike.

Figure 3(b) gives the complementary thermal constraint. The output mode must be radiatively overcoupled, and the pump port must address it weakly. For $\kappa_{fi}/\kappa_f \approx 1.9\%$, the additional pump heat remains below $10^{-2}\%$ for $\epsilon \lesssim 0.03$, and stays below $5 \times 10^{-2}\%$ over a much wider region. Conversely, a large internal-loss fraction rapidly removes the advantage even if the UCPL remains spectrally enhanced. The relevant non-Hermitian control variable is therefore not the total output linewidth by itself, but its partition into external

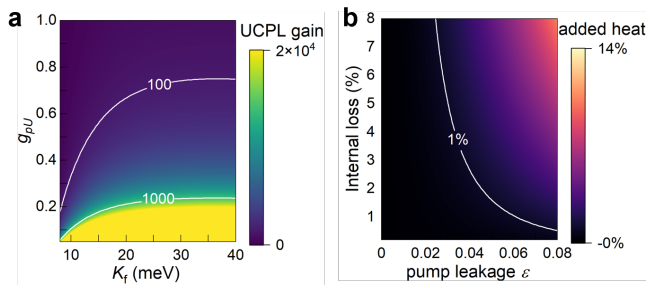

FIG. 3. Design maps for the high-selectivity regime. (a) UCPL enhancement at the EP as a function of the output linewidth κ_f and the undesired direct pump coupling g_{pU} to the upper radiative branch. (b) Added pump heat as a function of pump leakage ϵ and the internal-loss fraction κ_{fi}/κ_f of the output mode. The star marks the nominal design; white contours indicate representative gain and heating thresholds.

radiation and internal absorption. This distinction is important even in an all-dielectric implementation, because a broad linewidth may originate from either useful outcoupling or parasitic material and scattering loss. The maps identify two distinct failure modes. If g_{pU} approaches g_{fU}, the pump directly populates the upper radiative branch and the separation of optical functions is lost. If ϵ or κ_{fi}/κ_f is large, the output resonance contributes appreciably to pump heating even though the UCPL spectrum remains enhanced. These situations can produce a strong anti-Stokes sideband without improving the cooling balance. Useful operation therefore requires both modal selectivity and port selectivity. A separate no-retuning tolerance analysis is given in Supplemental Fig. S2. At the nominal coupling strength, more than 80% of the nominal EP-point UCPL response is retained for $|\Delta_f - \Delta_f^{\mathrm{EP}}| \lesssim 0.30\kappa_f$, and more than 50% remains within approximately $0.63\kappa_f$. Static inhomogeneous broadening of the upper radiative manifold reduces the absolute response continuously rather than abruptly: more than 80% of the zero-broadening EP UCPL remains for $\sigma_{\mathrm{inh}} \lesssim 0.38\gamma_U$, and more than 50% remains for $\sigma_{\mathrm{inh}} \lesssim \gamma_U$. Thus the mathematical EP is an exact degeneracy, but the experimentally useful high-response region has a finite tolerance.

We next connect the reduced model to the CdS cooling cycle and to the net-cooling criterion summarized in Fig. 4. CdS provides a conventional II–VI semiconductor mapping, with prior anti-Stokes upconversion and laser-cooling experiments under 514-nm excitation providing the material precedent [34, 35]. In the present model, L denotes an effective red-tail or localized near-band-edge excitation manifold populated by the 514-nm pump, whereas U denotes the radiative free-exciton or band-edge manifold emitting near 506 nm. The approximately 37-meV energy increment corresponds to one CdS LO phonon [36–38]. Figures 4(a)–4(c) quantify the emit-

ted spectrum and the fixed-power cooling window.

The mode-resolved heat ratio in Eq. (10) quantifies the additional loss introduced by the output branch, but it does not by itself include the material background absorption or the external quantum efficiency. To connect the optical enhancement to a device-level cooling criterion, we compare the architectures at fixed total absorbed pump power $P_{\rm abs}$ and write

$$\alpha_{\rm AS}^{(j)} = \mathcal{E}_j \alpha_{\rm AS}^{(0)}, \qquad \mathcal{E}_j = \{1, 344.99, 1226.59\}, \tag{12}$$

for the single-mode, off-EP dual-mode, and EP configurations, respectively. Here $\alpha_{\rm AS}^{(0)}$ is the effective useful anti-Stokes coefficient of the single-mode reference, while α_b collects absorption that produces heat without generating an escaping anti-Stokes photon. The fixed-power cooling efficiency is [12, 13]

$$\eta_{\rm cool}^{(j)} \equiv \frac{P_{\rm net}^{(j)}}{P_{\rm abs}} = \eta_{\rm ext} \frac{\omega_f}{\omega_p} \frac{\mathcal{E}_j \alpha_{\rm AS}^{(0)}}{\mathcal{E}_j \alpha_{\rm AS}^{(0)} + \alpha_b} - 1, \tag{13}$$

where $\eta_{\rm ext}$ is the external quantum efficiency of the phonon-assisted emission channel. Net cooling requires

$$\frac{\alpha_b}{\alpha_{\rm AS}^{(0)}} < \mathcal{E}_j \left(\eta_{\rm ext} \frac{\omega_f}{\omega_p} - 1 \right), \tag{14}$$

which is meaningful only when $\eta_{\rm ext} > \omega_p/\omega_f$. Figures 4(b) and 4(c) show the resulting net-cooling window for a 514-nm pump, $\hbar\omega_p = 2.412$ eV, and a 37-meV LO phonon. Thus $\omega_f/\omega_p = 1.01534$, and the zero-background condition requires $\eta_{\rm ext} > 0.98489$. At $\eta_{\rm ext} = 0.99$, the maximum tolerable background ratio increases from 5.19×10^{-3} in the single-mode reference to 1.79 in the off-EP dual-mode structure and 6.36 at the EP. The EP does not eliminate background absorption and does not relax the fundamental quantum-efficiency condition. It instead increases the useful anti-Stokes fraction of a fixed absorbed power, thereby enlarging the range of α_b over which the material can exhibit positive cooling power. As an illustrative scaling estimate, for $\eta_{\rm ext} = 0.99$ and $\alpha_b/\alpha_{\rm AS}^{(0)} = 1$, Eq. (13) gives $\eta_{\rm cool}^{\rm EP} \simeq 4.37 \times 10^{-3}$. An absorbed pump power of $P_{\rm abs} = 100$ μW would then correspond to $P_{\rm net} \simeq 0.44$ μW. The steady-state temperature reduction depends additionally on the thermal conductance $G_{\rm th}$ between the cooled CdS region and its environment, approximately $\Delta T \simeq P_{\rm net}/G_{\rm th}$. Because $G_{\rm th}$ depends on the dimensions of the cooled region, support geometry, substrate contact, and thermal boundary resistance, the present optical model does not assign a universal temperature drop. A partially suspended or weakly supported CdS structure would reduce $G_{\rm th}$ and allow this conversion to be calibrated experimentally. The theory suggests a direct experimental sequence. First, tune the 506-nm output resonance and its coupling to the radiative CdS manifold through the output EP while monitoring eigenfrequency splitting and linewidth exchange. Second, measure the total UCPL flux together with pump reflection, transmission, and absorbed power. Selective extraction predicts a strong UCPL maximum near the output EP with only a weak change in the pump response. Third, calibrate the external quantum efficiency and background absorption, and perform local thermometry to determine whether the operating point lies above the threshold in Fig. 4(c).

A cooling measurement should additionally distinguish increased extraction efficiency from a redistribution of the emission direction. The anti-Stokes photon flux must therefore be integrated over a calibrated radiation pattern rather than inferred from one collection angle. A useful control is to detune C_f while leaving C_p unchanged. The model predicts that UCPL follows the output eigenvalue topology, whereas absorbed pump power remains nearly flat. Observation of eigenvector coalescence, a coincident UCPL maximum, and positive local cooling would connect the passive EP directly to an improvement of the semiconductor refrigeration channel. A second requirement is spectral compatibility. The separation between C_p and C_f must track the material energy difference associated with the phonon-assisted transition. Finite linewidths tolerate modest mismatch, but a mismatch larger than the joint overlap of the reservoir, radiative state, and output resonance suppresses the product in Eq. (4). The scheme is therefore tied to a single microscopic cooling cycle rather than being a generic two-cavity intensity enhancement. The calculation is performed in the low-excitation linear-response regime, where the material decay rates are treated as effective constants. At higher excitation densities, surface recombination, exciton–exciton annihilation, Auger processes, state filling, and carrier- or temperature-induced broadening can make the effective nonradiative rate density and temperature dependent, schematically $\gamma_{\rm nr}(n,T) = \gamma_{\rm nr}^{(0)}(T) + A(T)n + B(T)n^2 + \cdots$. These processes will reduce the absolute UCPL and impose an upper bound on useful pump power. Importantly, the 1.23×10^3 system-level UCPL ratio is not a thousandfold increase in the pump-created carrier population: 344.99 of the enhancement arises from adding an independent extraction resonance and only a further factor of 3.56 from EP tuning, while the modeled pump-induced heat changes by only 2.8×10^{-3}%. The enhancement therefore primarily reflects improved conversion and escape of a phonon-promoted excitation rather than a comparable increase in carrier density. Nevertheless, nonlinear nonradiative pathways ultimately cap the usable enhancement, so the design maps should be interpreted as conditions on the low-excitation regime in which the upconversion response remains approximately linear and the external radiative efficiency remains high.

Having established the four-mode mechanism, its robustness, and the net-cooling requirements, we now ask

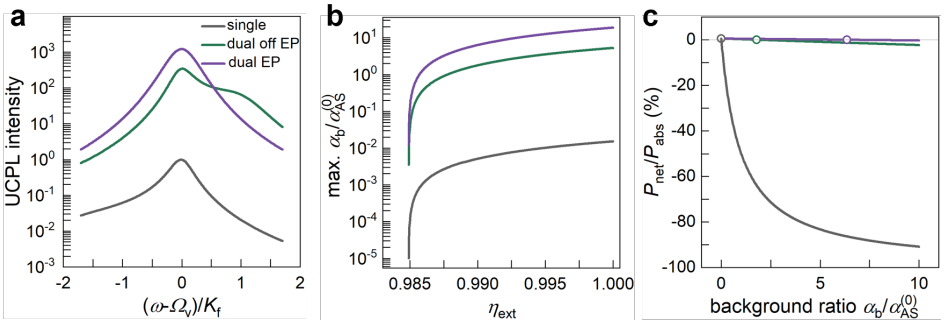

FIG. 4. CdS upconversion spectra and fixed-power net-cooling performance. (a) Calculated UCPL spectra for the single-mode, off-EP dual-mode, and EP configurations. (b) Maximum background-absorption ratio $\alpha_b/\alpha_{\mathrm{AS}}^{(0)}$ compatible with net cooling as a function of external quantum efficiency η_{ext}. The vertical dashed line marks the zero-background requirement $\eta_{\mathrm{ext}} = \omega_p/\omega_f$. (c) Cooling efficiency $P_{\mathrm{net}}/P_{\mathrm{abs}}$ versus background absorption at $\eta_{\mathrm{ext}} = 0.99$; open circles mark the zero-cooling thresholds.

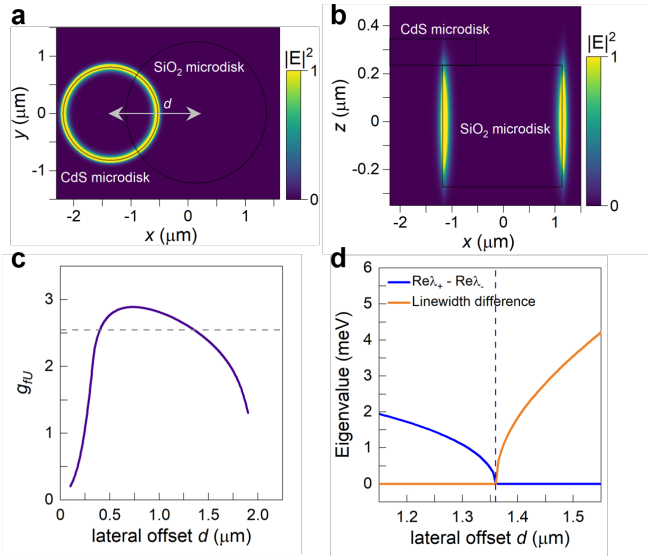

FIG. 5. Device-level microdisk realization and geometry-to-EP mapping. (a) Calculated 514-nm CdS pump WGM. (b) x–z distribution of the 506.23-nm SiO_2 output WGM, showing its evanescent overlap with the CdS disk. (c) Effective C_f–U coupling extracted from the three-dimensional overlap model as a function of lateral center offset d. The outer solution of $g_{fU} = |\kappa_f - \gamma_U|/4 = 2.545$ meV occurs at $d_{\mathrm{EP}} = 1.360$ μm, where $\Gamma_{\mathrm{CdS}} = 3.853 \times 10^{-3}$. (d) Real-frequency splitting and linewidth difference of the reduced C_f–U block, showing eigenvalue coalescence and linewidth exchange at the same geometric operating point.

whether the required optical functions can be realized in a concrete device. The device concept introduced schematically in Fig. 1(d) is analyzed quantitatively in Fig. 5. One feasible implementation is the vertically stacked, laterally offset double-microdisk shown in Fig. 5(a, b). The upper CdS microdisk acts as the 514-nm pump cavity C_p and is side-coupled to a dedicated input bus waveguide. The lower SiO_2 microdisk supports the approximately 506.23-nm output cavity C_f and is coupled to a separate output waveguide [39–45]. This arrangement separates pump injection from anti-Stokes extraction at the device level while allowing the evanescent tail of C_f to overlap the CdS active region. The lateral center offset d shown schematically in Fig. 1(d) is the same geometric control parameter used below to tune the C_f–U coupling through the mode overlap. The dimensions used for the feasibility estimate are $R_p = 0.834$ μm, $t_p = 100$ nm, $R_f = 1.167$ μm, and $t_f = 550$ nm.

The feasibility calculation uses an effective-index WGM overlap model [39, 46, 47]; the full derivation and numerical procedure are given in the Supplemental Material. For the dimensions in Fig. 5, the 514-nm CdS pump mode and 506.23-nm SiO_2 output mode are represented by the lowest vertical TE-like modes with azimuthal orders near $m_p = 21$ and $m_f = 20$. The output-mode evanescent field penetrates into the CdS disk [41, 47], and the resulting three-dimensional filling factor $\Gamma_{\mathrm{CdS}}(d)$ decreases continuously as the lateral offset is increased. Following standard cavity–exciton coupling models, we map this overlap to the calibrated light–matter coupling as $g_{fU}(d) = g_{\mathrm{bulk}}\sqrt{\Gamma_{\mathrm{CdS}}(d)}$ [48–50]. The outer crossing of the passive-EP condition $g_{fU} = |\kappa_f - \gamma_U|/4 = 2.545$ meV occurs at $d_{\mathrm{EP}} = 1.360$ μm, where $\Gamma_{\mathrm{CdS}} = 3.853 \times 10^{-3}$. Figure 5(d) shows that this same geometric point separates the frequency-split and linewidth-split sides of the C_f–U subsystem. Thus the reduced-model EP can be associated with a continuously tunable fabrication parameter rather than an abstract coupling constant. The coupling strengths to L and U are jointly determined by optical overlap, the CdS spectral response, and polarization selection rules, which are represented by the microscopic coupling parameters in the reduced model.

In conclusion, laser cooling of semiconductors is constrained by the fact that resonant enhancement often amplifies the useful and parasitic channels together. We remove this constraint by assigning reservoir preparation and anti-Stokes extraction to separate optical modes. The pump branch remains selective and weakly dissipa-

tive; the output branch is radiatively overcoupled and tuned to an EP with the upper state. The resulting EP-assisted modification is projected predominantly onto the final-state extraction channel rather than acting as a broadband field enhancement, while the dominant part of the total gain comes from separating preparation and extraction in the first place. For the representative CdS parameter set, the purely passive architecture produces a 1.23×10^3 UCPL enhancement, an optical selectivity gain of the same order, and only $2.8 \times 10^{-3}\%$ additional pump heating. At fixed absorbed power, this enhancement raises the tolerable background-absorption ratio by the same architecture-dependent factor, while the fundamental requirement $\eta_{\mathrm{ext}} > \omega_p/\omega_f$ remains unchanged. The physical prescription is explicit: separate the pump and anti-Stokes resonances, suppress pump leakage into the output mode, and ensure that the output linewidth is dominated by useful radiation. The microdisk construction further shows that the required output coupling can be mapped to a continuous geometric offset, providing a concrete route from the reduced non-Hermitian parameters to device tuning. This provides a general nanophotonic route toward local semiconductor refrigeration.

Acknowledgments. J. Z. acknowledges the funding support from National Natural Science Foundation of China (12525405), and the CAS Project for Young Scientists in Basic Research (YSBR-120).

Competing interests. The authors declare no competing interests.

Data availability. The authors declare that data generated in this study are provided in the paper. Further datasets are available from the corresponding author upon request.

* zhangjwill@semi.ac.cn

[1] Pringsheim P 1929 *Zeitschrift für Physik* **57** 739
[2] Chang H and Zhang J 2023 *Chip* **2** 100054
[3] Nemova G and Kashyap R 2010 *Rep. Prog. Phys.* **73** 086501
[4] Sheik-Bahae M and Epstein R I 2009 *Laser Photon. Rev.* **3** 67
[5] Epstein R I, Buchwald M I, Edwards B C, Gosnell T R, and Mungan C E 1995 *Nature* **377** 500
[6] Gosnell T R 1999 *Opt. Lett.* **24** 1041
[7] Seletskiy D V, Melgaard S D, Bigotta S, Di Lieto A, Tonelli M, and Sheik-Bahae M 2010 *Nat. Photonics* **4** 161
[8] Melgaard S D, Seletskiy D V, Di Lieto A, Tonelli M, and Sheik-Bahae M 2013 *Opt. Lett.* **38** 1588
[9] Melgaard S D, Albrecht A R, Hehlen M P, and Sheik-Bahae M 2016 *Sci. Rep.* **6** 20380
[10] Hehlen M P, Meng J, Albrecht A R, Lee E R, Gragossian A, Love S P, Hamilton C E, Epstein R I, and Sheik-Bahae M 2018 *Light Sci. Appl.* **7** 15
[11] Mobini E, Rostami S, Peysokhan M, Albrecht A, Kuhn S, Hein S, Hupel C, Nold J, Haarlammert N, Schreiber T, Eberhardt R, Tünnermann A, Sheik-Bahae M, and Mafi A 2020 *Commun. Phys.* **3** 134
[12] Sheik-Bahae M and Epstein R I 2007 *Nat. Photonics* **1** 693
[13] Melgaard S D, Seletskiy D V, Polyak V, Asmerom Y, and Sheik-Bahae M 2014 *Opt. Express* **22** 7756
[14] Rupper G, Kwong N H, and Binder R 2006 *Phys. Rev. Lett.* **97** 117401
[15] Li J Z 2007 *Phys. Rev. B* **75** 155315
[16] Khurgin J B 2007 *Phys. Rev. Lett.* **98** 177401
[17] Li D, Zhang J, and Xiong Q 2013 *Opt. Express* **21** 19302
[18] Ha S-T, Shen C, Zhang J, and Xiong Q 2016 *Nat. Photonics* **10** 115
[19] Pant A, Xia X, Davis E J, and Pauzauskie P J 2020 *Nat. Commun.* **11** 3235
[20] Benzaouia M and Fan S 2026 *Phys. Rev. Applied* **25** 014070
[21] Ju P, Shen K, Püschel S, Jin Y, Tanaka H, and Li T 2026 *Phys. Rev. Research* **8** 023149
[22] Mendicino L, Mayo F, Schmiegelow C, and Roncaglia A 2025 *Phys. Rev. A* **112** 063522
[23] Chen Y-C and Bahl G 2015 *Optica* **2** 893
[24] Berry M V 2004 *Czech. J. Phys.* **54** 1039
[25] Rotter I 2009 *J. Phys. A* **42** 153001
[26] Heiss W D 2012 *J. Phys. A* **45** 444016
[27] Miri M-A and Alù A 2019 *Science* **363** eaar7709
[28] Özdemir Ş K, Rotter S, Nori F, and Yang L 2019 *Nat. Mater.* **18** 783
[29] Lin Z, Ramezani H, Eichelkraut T, Kottos T, Cao H, and Christodoulides D N 2011 *Phys. Rev. Lett.* **106** 213901
[30] Pick A, Zhen B, Miller O D, Hsu C W, Hernandez F, Rodriguez A W, Soljačić M, and Johnson S G 2017 *Opt. Express* **25** 12325
[31] Langbein W 2018 *Phys. Rev. A* **98** 023805
[32] Ferrier L, Bouteyre P, Pick A, Cueff S, Dang N H M, Diederichs C, Belarouci A, Benyattou T, Zhao J X, Su R, Xing J, Xiong Q, and Nguyen H S 2022 *Phys. Rev. Lett.* **129** 083602
[33] Lau H-K and Clerk A A 2018 *Nat. Commun.* **9** 4320
[34] Zhang J, Li D, Chen R, and Xiong Q 2013 *Nature* **493** 504
[35] Li D, Zhang J, Wang X, Huang B, and Xiong Q 2014 *Nano Lett.* **14** 4724
[36] Khurgin J B 2008 *Phys. Rev. B* **77** 235206
[37] Li D, Zhang J, and Xiong Q 2012 *ACS Nano* **6** 5283
[38] Xu X, Zhao Y, Sie E J, Lu Y, Liu B, Ekahana S A, Ju X, Jiang Q, Wang J, and Xiong Q 2011 *ACS Nano* **5** 3660
[39] McCall S L, Levi A F J, Slusher R E, Pearton S J, and Logan R A 1992 *Appl. Phys. Lett.* **60** 289
[40] Little B E, Chu S T, Haus H A, Foresi J, and Laine J-P 1997 *J. Lightwave Technol.* **15** 998
[41] Kippenberg T J, Spillane S M, Armani D K, and Vahala K J 2003 *Appl. Phys. Lett.* **83** 797
[42] Vahala K J 2003 *Nature* **424** 839
[43] Ilchenko V S and Matsko A B 2006 *IEEE J. Sel. Top. Quantum Electron.* **12** 15
[44] Gorodetsky M L and Ilchenko V S 1999 *J. Opt. Soc. Am. B* **16** 147
[45] Yariv A 2002 *IEEE Photon. Technol. Lett.* **14** 483
[46] Chin M K, Chu D Y, and Ho S-T 1994 *Opt. Commun.* **109** 467
[47] Rosenblit M, Horak P, Helsby S, and Folman R 2004 *Phys. Rev. A* **70** 053808
[48] Weisbuch C, Nishioka M, Ishikawa A, and Arakawa Y

1992 *Phys. Rev. Lett.* **69** 3314
[49] Ajiki H, Kaneno T, and Ishihara H 2006 *Phys. Rev. B* **73** 155322
[50] Khitrova G, Gibbs H M, Kira M, Koch S W, and Scherer A 2006 *Nat. Phys.* **2** 81