



Tunable bandgap renormalization by nonlocal ultra-strong coupling in nanophotonics

Yaniv Kurman and Ido Kaminer

In quantum optics, great effort is being invested in enhancing the interaction of quantum emitters with light. The different approaches include increasing the number of emitters, the laser intensity or the local photonic density of states at the location of an atom-like localized emitter. In contrast, solid-state extended emitters hold an unappreciated promise of vastly greater enhancements through their large number of vacant electronic valence states. However, the majority of valence states are considered optically inaccessible by a conduction electron. We show that, by interfacing three-dimensional (3D) solids with 2D materials, we can unlock the unoccupied valence states by nonlocal optical interactions that lead to ultra-strong coupling for each conduction electron. Consequently, nonlocal optical interactions fundamentally alter the role of the quantum vacuum in solids and create a new type of tunable mass renormalization and bandgap renormalization, which reach tens of millielectronvolts in the example we show. To present quantitative predictions, we develop a non-perturbative macroscopic quantum electrodynamic formalism that we demonstrate on a graphene-semiconductor-metal nanostructure. We find new effects, such as nonlocal Rabi oscillations and femtosecond-scale optical relaxation, overcoming all other solid relaxation mechanisms and fundamentally altering the role of optical interactions in solids.

Electronic systems in both condensed matter and atomic physics share an inherent similarity: their electrons follow the same basic rules for optical interaction. It is no coincidence that in both solids and gases, a typical emission timescale for an excited electron is on the order of nanoseconds^{1,2}. The key to this universal behaviour is in the intrinsic transition matrix elements governing the interaction of a charge carrier with light, whereby a similar order of magnitude is acquired in solid and atomic systems. Additionally, because the total emission rate for each electron depends also on the photonic density of states^{3,4}, it is prevalent in both systems to enhance emission rates by optical environment engineering, that is, the Purcell factor^{5,6}. Independently, an increase in the number of emitters sharing the same optical mode enhances the total coupling by a coherent collective response in extended solid systems as quantum wells (QWs)^{7–10} and in an ensemble of localized atom-like emitters (for example, molecules)^{11,12}. Such systems can achieve a collective ultra-strong coupling (USC) thanks to the large number of emitters. All these techniques enable the observation of Rabi oscillations in the strong coupling regime, which occur when the coupling frequency overcomes the relaxation rates, and in the USC regime, which occur when the coupling frequency reaches the transition frequency^{13–15}.

However, there is one critical aspect in which solid-state and atom-like systems differ significantly. Unlike localized atom-like emitters, extended solid-state emitters may have a high number of electronic final states available for each electron, apparently promising extreme enhancements of the total emission and rates. Why then is the typical emission of each excited electron in semiconductors still limited to gigahertz rates (nanosecond lifetimes), as in most natural and artificial atoms? Why are the leading (dipole) optical emission rates in extended solids comparable to the emission rates in localized atoms, ignoring the potentially many valence band final states?

There are two reasons why extended solid systems are constrained from reaching far faster light–matter interaction rates:

(1) the conservation of energy limits the number of accessible electronic final states, fixed by the emitted photon energy (Fig. 1a) and (2) even at ultra-short timescales, during which the conservation of energy is not enforced, the conservation of momentum still enables only a small number of accessible electronic final states because of the momentum mismatch between the photon and the charge carrier (Fig. 1b). Consequently, despite the existence of many more available final electronic states in the valence band, the number of accessible states is strictly limited for optical interaction and the vast majority of available valence states remain prohibited.

Our work shows how these limiting factors can be bypassed by using nonlocal (momentum-dependent) optical interactions in nanophotonic structures. In particular, interfacing solids with two-dimensional (2D) materials that support optical excitations, that is, polaritons, can enable transitions to the valence states that are conventionally prohibited for optical transitions but become accessible by the polaritons' nonlocal interactions.

A polariton, a hybrid of a photon and a material excitation, takes the role of a photon in the light–matter matrix element^{16–18}. In nanophotonic structures and settings of 2D materials, polaritons create novel light–matter interactions because their momentum can be increased by more than two orders of magnitude relative to the free-space photon at the same frequency^{19–21}. This increased polariton momentum is especially influential for interactions with quantum emitters, because it extremely enhances the photonic density of states^{22–24} and enables interactions that are typically forbidden^{25–27}. Specifically, in solid quantum emitters, such as QWs, nearby propagating polaritons enable non-vertical (momentum-dependent) electron transitions that create a nonlocal optical response for the structure as a whole²⁶. The nonlocal optical response necessitates a spatial dispersion in the material's permittivity, which can no longer be a function solely of the optical frequency and a single location variable. A pronounced signature of this nonlocality is found in nanophotonic structures, such as graphene sheets^{26,28}, especially small nanoparticles²⁹ and various metamaterials³⁰.

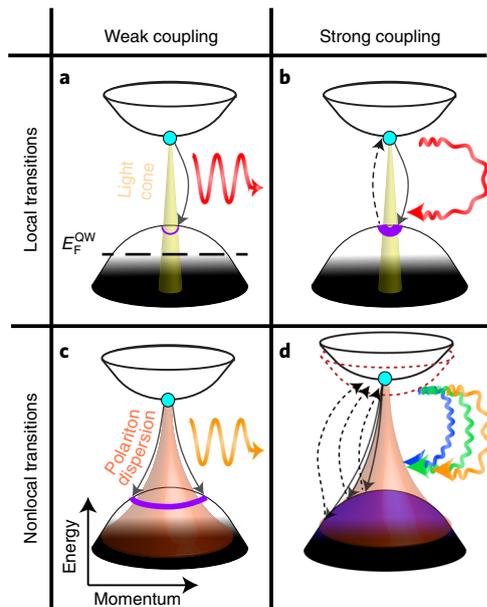


Fig. 1 | The role of nonlocality in reaching USC in solids. a–d. A general scheme depicting weak coupling (**a,c**) and strong coupling (**b,d**) for local (vertical) interaction with photons (light cone in yellow) (**a,b**) and for nonlocal (non-vertical) interaction with polaritons (polariton dispersion in pink) (**c,d**). In weak coupling, the photon or polariton is spontaneously emitted at one specific energy according to energy conservation, whereas in strong coupling, the emitted photon or polariton can be emitted and reabsorbed even when off-resonant. **c**, The large momenta of the polaritons relative to the small momenta of the photons enable nonlocal interactions and thus open coupling to a wider range of final states (depicted in purple) for the excited electron (turquoise). **d**, Provided sufficiently strong coupling and nonlocal interactions, a single excited electron can couple to all available final states, even if detuned from the polaritonic dispersion. This coupling is still limited by the final state occupation according to the Fermi–Dirac distribution (unoccupied and occupied states in black and white, respectively). Consequently, these enhanced nonlocal interactions cause an electron–polariton renormalization of the semiconductor bandgap and the carrier effective mass, both of which can be externally tuned by controlling the polaritonic properties or the Fermi energy.

In Fig. 1 we illustrate how nonlocality can be utilized for reaching USC. For weak coupling interactions (Fig. 1c), the nonlocal emission enables non-vertical interband emission, which increases the number of accessible final electronic states (purple area) and thus increases the total emission rate. Moreover, the number of accessible final electronic states can be further increased by enhancing off-resonance optical interaction so they may contribute over short timescales. As a result, we find that polaritons with an extreme density of states enable each individual conduction electron to interact with all available final (valence band) states as in a continuum of cavity polaritonic modes (Fig. 1d). Consequently, we show that a single conduction electron can reach Rabi oscillations with polaritonic vacuum fluctuations. The oscillation rate may overcome the rate of loss by other mechanisms (that is, strong coupling) and even become comparable to the transition frequency (that is, USC). This electron–polariton USC in vacuum creates a novel kind of mass and bandgap renormalization (shifted conduction curve, Fig. 1d) that can be externally tuned by controlling the number of available valence states or the optical environment, that is, the polaritons’ properties. From a fundamental perspective, this makes the solid’s optical interaction surprisingly similar to an electron–phonon interaction.

A qualitative comparison of electron–phonon interactions in solids and our electron–polariton counterpart could highlight the consequences of the novel nonlocal USC interactions. Electron–phonon interactions are well known to have relaxation rates far higher than regular optical emission rates (subpicosecond versus nanosecond)^{2,31}. The root cause for this extreme difference in rates can be explained by the extended nature of electron wavefunctions in solids, with a wavelength larger than the phononic wavelength but considerably smaller than the photonic wavelength. This inequality makes the electrons typically heavy as compared to photons and light as compared to phonons. Consequently, when interacting with the phonons, electrons cannot be regarded as point particles, as in the commonly used dipole approximation for optical interactions. Instead, the electron–phonon interaction is always nonlocal—it requires a spatial overlap with the coherently moving phonon wave^{32,33}. Thus, phononic and nonlocal optical interactions can cause significant changes in the electron wavefunction, which results in a larger number of accessible electronic final states as compared to regular optical interaction and, consequently, results in higher relaxation rates. Thus, we show how polaritonic emission, by providing access to many electronic final states, becomes a competitor for dominating the charge carrier relaxation and thus overturning the role of optical relaxation in solids.

Here, we find that the nonlocal nature of electron–polariton interactions opens up USC effects at optical frequencies at the level of each conduction electron in the solid. As a vacuum effect, the USC causes mass and bandgap renormalization reaching 100 meV, similar to the vacuum Rabi splitting in atomic systems. To describe these novel electron–polariton interactions quantitatively, we use an extended Wigner–Weisskopf (WW) model to develop a non-perturbative macroscopic quantum electrodynamics (QED) formalism. Our new formalism captures the nonlocal interaction of polaritons with the entire electron wavefunction of the solid quantum emitters, rather than a point emitter, and tracks the non-perturbative interaction dynamics. We exemplify our predicted effects on a 5 nm InAs/InP QW, situated between a graphene sheet and a perfect conductor (for example, gold), which forces an extreme density of polaritonic states²⁴ and confines the light inside the semiconductor volume (Fig. 2a). Our nonlocal WW formalism shows that polariton emission dominates the relaxation dynamics with timescales of tens of femtoseconds. Such a relaxation includes nonlocal Rabi oscillations at these timescales, reaching the USC regime. Our formalism uncovers the large degree of tunability in our quantitative results by altering the number of unoccupied final states (holes) in the semiconductor or by changing the external gate-tunable graphene Fermi energy to control the polaritonic dispersion.

Results

Theory of nonlocal USC. In this section, we present the main steps in our nonlocal WW formalism, using macroscopic QED theory^{34,35}. This new formalism extends beyond the atomic and single-mode WW theory, and it allows us to capture the full time dynamics for the interaction between an arbitrary quantum emitter and an arbitrary optical environment with or without losses and with or without nonlocality. Similar approaches were originally developed for localized atomic emitters^{13,36} and we now show how they can be extended to condensed-matter systems and along the way also retrieve the formula for bandgap renormalization.

As a first step, we treat the quantum state as a superposition of all possible final states ($|f\rangle$) with all possible single quanta of optical excitations (M) and a single initial excited state ($|i\rangle$) with 0 optical excitations:

$$\begin{aligned}
 |\psi(t)\rangle &= C_i(t)e^{-i\omega_i t}|i\rangle \otimes |0, \dots, 0\rangle \\
 &+ \sum_f \sum_M C_{f,M}(t)e^{-i(\omega_f + \omega_M)t}|f\rangle \\
 &\otimes |0, \dots, 0, 1_M, 0, \dots, 0\rangle
 \end{aligned} \quad (1)$$

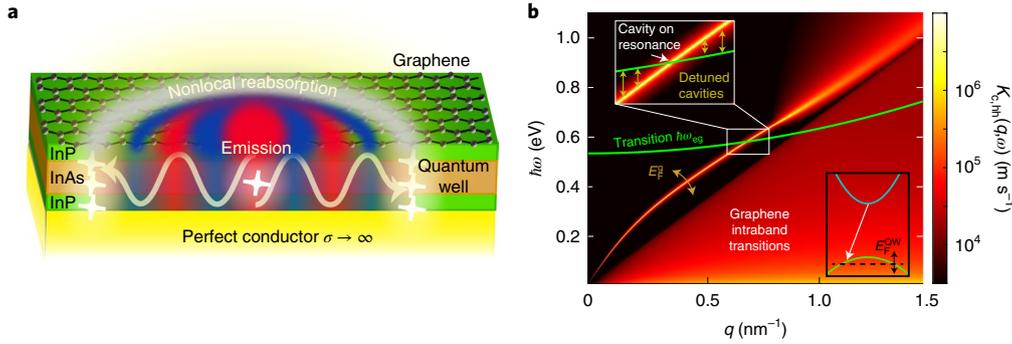


Fig. 2 | The ultra-strongly coupled nanophotonic system description. **a**, Illustration of the graphene–semiconductor–metal structure—an InAs/InP QW (orange/green, respectively) of width 5 nm, with an active layer of 4 nm and bandgap of 0.53 eV, stated between a graphene sheet and a perfect conductor (for example, gold). The intuition behind the nonlocal USC is the possibility that large-momentum graphene plasmons (red and blue) can be emitted in one spot and reabsorbed (white stars) in other locations by the same extended electron wavefunction. **b**, The coupling strength between the conduction to heavy-hole transitions in the QW and each optical excitation (polariton) with specific in-plane momenta \mathbf{q} and frequency ω , depicted by the kernel $K_{c,hh}(\mathbf{q}, \omega)$, which governs the optical dynamics (equation (4)). Specifically, the kernel shows the plasmon’s dispersion (tuned by the graphene Fermi level E_F^W) alongside the graphene intraband transitions. The dynamics is equivalent to a continuum of cavity modes (independent for each \mathbf{q}), all coupling to the same conduction electron with a different energy detuning for each cavity. Top inset: details of the on-resonance cavity mode environment. Bottom inset: illustration of the QW band structure (see Supplementary Fig. 1 for more details). The plot assumes one excited electron at the conduction band’s lowest energy. The valence occupation can be controlled through the QW Fermi level E_F^{QW} .

Here, $C_{i,M}(t)$ and $C_i(t)$ are the complex amplitudes, so their square moduli will be the probabilities. $\hbar\omega_M$ is the optical excitation energy and $\hbar\omega_i$ ($\hbar\omega_f$) is the electron energy at the initial (final) state. For example, in the case of a semiconductor interband emission with a direct bandgap E_{gap} , $\hbar\omega_i = E_{gap} + \frac{\hbar^2 k_i^2}{2m_e^*}$ and $\hbar\omega_f = -\frac{\hbar^2 k_f^2}{2m_h^*}$, with m_e^* and m_h^* being the effective masses (for example, $m_f^* = m_{hh}^*$ for heavy holes) and $\mathbf{k}_{i,f}$ their in-plane momenta. Both summations in equation (1) can be discrete or continuous and this formalism could thus be reduced to the discrete multimode strong coupling as in superconducting systems³⁷, or a single-mode case as in the Jaynes–Cummings model.

The system’s evolution in time is determined by the interaction Hamiltonian, which can be written in two equivalent forms:

$$H_{\text{int}} = \frac{e}{m_0} \hat{\mathbf{p}} \cdot \hat{\mathbf{A}} = \sum_{\mathbf{f}} \sum_M \hbar g_{i,fM} \hat{b}_i^\dagger \hat{b}_f \hat{a}_M + \text{h.c.} \quad (2)$$

first with the momentum ($\hat{\mathbf{p}}$) and vector potential ($\hat{\mathbf{A}}$) operators (including the electron’s rest mass m_0 and charge e) and second with the creation and annihilation operators for the optical excitation, \hat{a}_M^\dagger and \hat{a}_M , and for the initial (final) electron states, \hat{b}_i^\dagger (\hat{b}_f^\dagger) and \hat{b}_i (\hat{b}_f). The interaction strength is given by the coupling constant $g_{i,fM}$, between a specific initial excited state, a final state and an optical excitation.

The determination of each coupling constant for general optical excitations in general optical environments can be remarkably challenging. In the macroscopic QED formalism, the vector potential is fully described through the dyadic Green’s function, $G(\mathbf{r}, \mathbf{r}'; \omega)$, which contains the electrical response from an electric dipolar excitation in any general (lossy) media. Using this formalism, we prove for any general electronic system and optical environment (see Methods):

$$\sum_M |g_{i,fM}|^2 e^{-i(\omega_M - \omega_{if})\tau} = \frac{4\pi}{m_0^2 c} \int d\omega e^{-i(\omega - \omega_{if})\tau} \int d\mathbf{r} \int d\mathbf{r}' \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \text{Im} \left[\overleftrightarrow{G}_{\text{lm}}(\mathbf{r}, \mathbf{r}'; \omega) \right] (p_l \psi_f(\mathbf{r})) (p_m \psi_f^*(\mathbf{r}')) \quad (3)$$

where $\psi_{i,f}$ represents the initial and final electron wavefunctions, α is the fine-structure constant, c is the speed of light in vacuum, and $\omega_{if} = \omega_i - \omega_f$. Here, $\hbar(\omega - \omega_{if})$ is the detuning energy that causes a destructive interference that gradually diminishes the coherence and leads to an overall decay as time (τ) evolves.

In a general layered structure, each optical excitation is well described by a specific frequency ω and in-plane wavevector \mathbf{q} . Hence, we can define a coupling kernel $K(\omega, \mathbf{q})$ as the spatial Fourier transform of the integrand in equation (3) that captures all the details of the coupling between the optical and electronic environments (Fig. 2b). In the special case of an interband emission in a graphene–semiconductor–metal structure of width d , each spectral component of equation (3) gets an insightful form of $\int d\mathbf{q} \text{Im} \left[\frac{r_{\text{gr}} e^{-2qd} - 1}{r_{\text{gr}} e^{-2qd} + 1} \right] |M_{c,hh}|^2 \delta_{\mathbf{k}_i, \mathbf{k}_f + \mathbf{q}}$. This expression includes the conduction to heavy-hole matrix element $|M_{c,hh}|^2$, the surface plasmon dispersion through $\text{Im} \left[\frac{r_{\text{gr}} e^{-2qd} - 1}{r_{\text{gr}} e^{-2qd} + 1} \right]$ where r_{gr} is the graphene Fresnel coefficient, and the conservation of in-plane momentum term $\delta_{\mathbf{k}_i, \mathbf{k}_f + \mathbf{q}}$ (Supplementary Section 1). Conservation of the in-plane momentum is a compelling outcome of the structure’s translational symmetry, because it also helps emphasize the importance of the nonlocal coupling arising from the electron extended wavefunction (rather than using the dipole approximation discussed in Supplementary Section 5). Indeed, Fig. 2b shows that the strongest coupling occurs along the graphene plasmons’ dispersion, and yet other graphene excitations (that is, the graphene particle–hole continuum^{38,39}) also contribute.

The complete dynamics of the wavefunction (equation (1)) follows the Schrödinger equation, into which we substitute equation (3) to yield an integro-differential equation for the initial excited state amplitude (more details are provided in Supplementary Section 1),

$$\dot{C}_i(t) = - \int_0^t dt' \sum_{\mathbf{f}} \sum_M |g_{i,fM}|^2 e^{-i(\omega_f + \omega_M - \omega_i)(t-t')} C_i(t') \quad \text{in 2D translational symmetric systems} \quad \Rightarrow$$

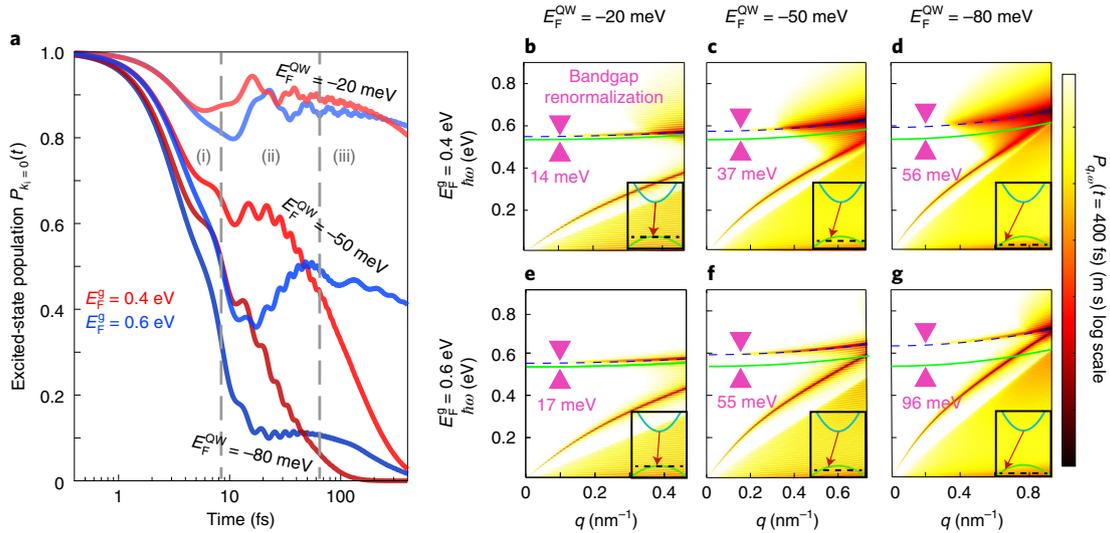


Fig. 3 | Controlling bandgap renormalization and reaching USC by altering the Fermi levels. **a**, The time-dependent initial state population probability $P_{k_i=0}(t)$ for different values of Fermi levels of both the graphene, E_F^g , and the QW, E_F^{QW} . We show the three decay stages. (i) An extremely fast initial decay followed by (ii) Rabi oscillations, which are evidence of non-perturbative phenomena, and eventually (iii) an exponential decay. **b–g**, The optical excitations' occupation probabilities $P_{q,\omega}$ for $E_F^g = 0.4$ eV (**b–d**) and 0.6 eV (**e–g**) when $E_F^{QW} = -20$ meV (**b,e**), -50 meV (**c,f**) and -80 meV (**d,g**). In the weak coupling regime, the optical excitations should follow the original band structure (green curve) because of conservation of energy and momentum. However, in the USC regime, the optical excitations follow the original band structure but in shifted energies (dashed blue curve). This shift is the bandgap renormalization (pink triangles), which also matches our direct prediction through the non-perturbative calculation of the energy shift (analogous to a Lamb shift). The plots are made at $t = 400$ fs to demonstrate both the emission into the shifted energy levels at sufficiently long times, together with a remaining occupation along the plasmon dispersion at sufficiently short times. More details about the system's time evolution, including the initial rapid decay, are provided in the Supplementary Video.

$$\dot{C}_{k_i}(t) = - \int_0^t dt' \int d\omega \int d\mathbf{q} K(\mathbf{q}, \omega) [1 - f_{F,D}(\omega_{k_f}(\mathbf{q}))] e^{-i(\omega - \omega_{k_i}(\mathbf{q}))(t-t')} C_{k_i}(t') \quad (4)$$

We introduce the zero-temperature Fermi–Dirac distribution ($f_{F,D}$) to place the Fermi level as a cutoff for the available final states; nevertheless, our qualitative predictions remain correct for room temperature. When comparing equation (4) to a single-mode cavity QED model, our system can be thought of as an interaction between a single electron/emitter and a continuum of cavities with a different energy detuning per cavity \mathbf{q} .

Our formalism assumes up to a single optical excitation that couples to each electron–hole path, similar to the rotating wave approximation (RWA). We also extend our formalism beyond the RWA and provide an analytical model to analyse the counter-rotating terms, thus converting our model into an extended Rabi-type model, yet with a continuum of optical modes. This analysis shows rigorously that the RWA is a good approximation for the parameters in our work (Supplementary Section 7 and Supplementary Fig. 5). Using macroscopic QED in our WW formalism inherently captures optical losses (for example, Landau damping) and we generalize it even further to include plasmon reabsorption by other valence electrons (Supplementary Section 8). In addition, our formalism may also capture additional external energy relaxation mechanisms (for example, electron–phonon and electron–electron scattering) that can be included in a phenomenological manner (Supplementary Section 3); these mechanisms add a decoherence rate that partially dampens the dynamics (Supplementary Fig. 4).

Controlling internal semiconductor properties by creating nonlocal USC. In this section, we use our theory above to show how large-momentum polaritons such as graphene plasmons can tune

the internal electronic properties of a semiconductor when their interaction reaches USC. The structure chosen as the chief example in our work is a QW (InAs, for more details see Supplementary Fig. 1) that contains extended (nonlocal in space) electron wavefunctions, as any solid, but has a small and direct bandgap. The optical environment supports large-momentum polaritons with extreme density of states near the bandgap energies.

Using our nonlocal WW formalism, we present the main phenomena related to the nonlocal USC in Fig. 3. Figure 3a shows the unique time dynamics of the initial excited-state occupation probability, $P_{k_i=0}(t) = |C_{k_i=0}(t)|^2$, while in Fig. 3b–g we present the occupation probability of each optical excitation, $P_{q,\omega}(t)$, long after the beginning of the emission process. This probability is calculated as

$$P_{q,\omega}(t) = |C_{q,\omega}(t)|^2 = K_{c,hh}(\mathbf{q}, \omega) \left| -i \int_0^t dt' e^{i(\omega - \omega_{k_i}(\mathbf{q}))t'} C_{k_i}(t') \right|^2 (1 - f_{F,D}(\omega_{k_f}(\mathbf{q}))) \quad (5)$$

which satisfies $P_{k_i}(t) + \int d\omega \int d\mathbf{q} P_{q,\omega}(t) = 1$ at all times t . Plotting equation (5) in Fig. 3b–g alongside equation (4) in Fig. 3a (and the Supplementary Video) reveals ultra-fast relaxation, the USC regime and bandgap renormalization during three qualitatively distinct stages in the excited-state decay:

(1) An ultra-fast optical relaxation of the electron occupation probability in less than 10 fs. Such a rapid decay is caused by the coupling into all accessible optical excitation modes (\mathbf{q}, ω) before conservation of energy is enforced. As a result, we find in Fig. 3b–g a non-zero occupation probability along the graphene plasmon dispersion due to the highest density of states. Such a short relaxation time proves the dominance of nonlocal electron–polariton relaxation over competing processes near the bottom of the conduction band in semiconductors⁴⁰.

(2) The initial rapid decay converts to Rabi oscillations with cycle below 10 fs, which indicates the USC regime, as the ratio between the Rabi oscillation frequency and the emission frequencies is ~ 1 . These oscillations involve fewer low-loss dominant modes compared to the many more modes involved in the initial decay. Nevertheless, the multimode nature of the process cause negligible counter-rotating contributions (Supplementary Fig. 5) and maintains the Rabi oscillations despite competing decoherence mechanisms and damping by the QW excited electron and other valence electrons (Supplementary Fig. 6).

(3) As time evolves, the Rabi oscillations fade out, and the optical occupation increases only for the excitations that conserve energy and momentum according to the QW's electronic band structure. This emission process is expected to occupy the excitations along the green curve in Figs. 2b and 3b–g. However, most of the occupied optical excitations are obtained along a shifted curve (dashed blue) that goes exactly parallel to the green curve, indicating an energy shift for the initial excited state, that is, a bandgap renormalization. Accordingly, the bandgap renormalization determines the relaxation dynamics at this stage and for long times. For example, when $E_F^{QW} = -50$ meV and $E_F^g = 0.6$ eV, the decay rate is slow (middle blue curve, Fig. 3a) because of the lack of polaritonic density of states along the renormalized band structure (no crossing in Fig. 3f). However, when changing E_F^g to 0.4 eV (Fig. 3a), a much faster decay rate is found because the coupling to the graphene particle-hole continuum becomes possible (Fig. 3c).

We note two control knobs that enable tuning of the bandgap renormalization and the dynamics of the system: varying the graphene Fermi level E_F^g (examples of 0.4 and 0.6 eV are shown in the top and bottom rows of Fig. 3b–g) and varying E_F^{QW} (values between -20 and -80 meV are shown in the columns of Fig. 3b–g, corresponding to hole concentrations of 0.4 to $1.7 \times 10^{19} \text{ cm}^{-3}$). As a result, a plot of $P_{q,\omega}(t)$ reveals tunable bandgap renormalization of tens of meVs, extremely large compared to the original bandgap (~ 530 meV). For comparison, we apply our formalism on the intersubband optical transition (Supplementary Fig. 3), where the weak coupling does not yield a significant energy shift, and the occupied optical excitations $P_{q,\omega}(t)$ follow the expected curve.

The nonlocal USC occurs on the level of each individual excited electron in the conduction band. As opposed to all previous examples of vacuum Rabi splitting (or Lamb shift) in solid-state systems, which depend on a high density of excitations (often modelled as two-level systems¹⁰), the nonlocal USC enable a single electron to interact with all the valence states. This possibility causes a significant energy shift in each single electron excitation, without requiring charge density in the conduction band. Consequently, we find the new effect of an effective mass renormalization, as shown in Fig. 4a. Intuitively, non-zero values of the excited-state in-plane momenta \mathbf{k}_i break the azimuthal symmetry in the emission so that the coupling to polaritons that are emitted in the \mathbf{k}_i direction will be reduced because of the lack of available valence states. Consequently, the excited energy level shift, $\delta\omega_{\mathbf{k}_i}$, is reduced when the excited state's momenta \mathbf{k}_i are increased, leading to flattening of the conduction band and an increase in the effective mass.

We extract an explicit formula for the energy shift that converges to a form similar to the electron-phonon bandgap renormalization³³:

$$\delta\omega_{\mathbf{k}_i} = \mathcal{P} \int d\omega \int d\mathbf{q} \frac{K_{c,hh}(\mathbf{q},\omega) [1 - f_{FD}(\omega_{\mathbf{k}_f}(\mathbf{q}))]}{\omega - \omega_{\mathbf{k}_i}(\mathbf{q}) - \delta\omega_{\mathbf{k}_i}} \quad (6)$$

where \mathcal{P} denotes the principal value. Equation (6) gives the real part of the electron self-energy, expressed through the kernel $K_{c,hh}(\mathbf{q}, \omega)$ and integrated over the entire space of optical excitations (\mathbf{q}, ω) . Thus, all momentum-conserving optical excitations are involved in the bandgap renormalization as in the initial ultra-fast decay.

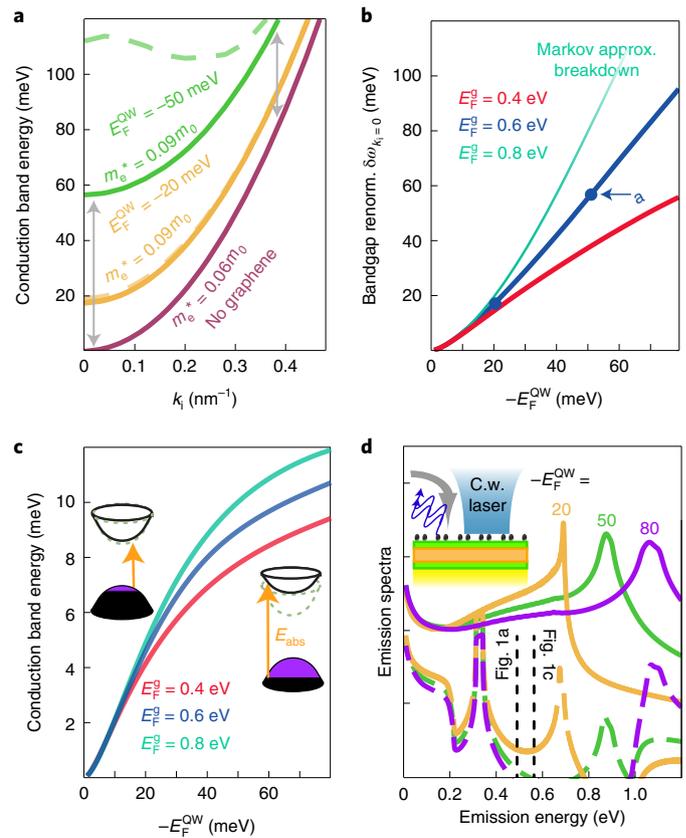


Fig. 4 | Tunable non-perturbative phenomena in electron-polariton USC.

a, Tunable mass and optical bandgap renormalization due to electron-polariton USC when $E_F^g = 0.6$ eV. Electrons with higher momentum exhibit smaller energy shifts (grey double arrows), leading to a change in the conduction electron effective mass. The dashed curves are calculated using only the second-order perturbation theory. **b**, Tunable bandgap renormalization as a function of E_F^{QW} for three different E_F^g values. When the doping of both the graphene and the QW is high, the Markov approximation in equation (6) fails; the renormalization can still be extracted from $P_{q,\omega}$, as in Fig. 3, which does not use this approximation. **c**, Shift in the photon absorption spectrum due to the bandgap and effective mass renormalization (increase in the minimal energy of absorption). Insets: the minimal-energy photon absorption (orange) by the most energetic valence electron (empty states in purple) into the renormalized conduction band (the original conduction band is shown by dashed lines). **d**, The highly tunable expected near-field emission spectra as a result of a continuous-wave (c.w.) laser excitation when $E_F^g = 0.6$ eV, and $E_F^{QW} = -20, -50$ and -80 meV (yellow, green and purple, respectively). Due to the enhanced optical response, the excited electron would prefer to emit into the near-field optical excitations, before a quasi-thermal equilibrium is reached in the conduction band. To couple out the near-field emission one can consider a metallic tip (solid) or a 20 nm periodic grating (dashed). Inset: illustration of the suggested experimental scheme.

This explicit formula requires the use of the Markov approximation (Supplementary Section 2), which is found to give an excellent match with the exact calculation through the optical occupation probability (equation (5) and Fig. 3b–g). Only at extreme values of doping (partially transparent curve ending at the top of Fig. 4b) does the Markov approximation cease to be valid; however, in such cases one also has to consider additional corrections, such as the second heavy-hole subband and the light-hole subbands (Supplementary Fig. 1). Using equation (6), we can directly calculate the bandgap renormalization $\delta\omega_{\mathbf{k}_i=0}$ and the mass renormalization, highlighting

their large degree of tunability (Fig. 4a,b). We stress that equation (6) uses all perturbation orders, instead of stopping in the second-order QED (as is the case for the Lamb shift in atomic physics), which makes a difference when increasing the hole population (dashed curves in Fig. 4a).

Discussion

To achieve the predicted nonlocal USC phenomena, a structure must support an extreme polaritonic density of states over a range of large-momentum polaritonic modes. This scheme enables a single electron to couple to more optical excitations than in conventional optical transitions and to decay to more states. Consequently, the nonlocal USC can be induced in 2D or thin 3D semiconductors when interfaced with 2D materials or metamaterials that support large-momentum polaritons (for example, graphene nanoribbons, hexagonal boron nitride, transition metal dichalcogenides or hyperbolic structures^{20,41–44}). None of the qualitative predictions depends on the properties of the specific solid or polaritonic interaction. Both interband and intersubband transitions are conceptually eligible to lead to the effects. We have focused discussion on interband emission because this also causes bandgap renormalization. We expect similarly strong or even stronger effects for bandgap renormalization at higher bands, as they could have more empty states with which to interact (although faster competitive hot carrier relaxation times⁴⁵).

Several schemes could be considered for experimental observation of our predictions. The bandgap renormalization can be observed experimentally with an electrical differential conductance measurement, dI/dV , in the QW. Nonlocal Rabi oscillations might be observed with ultra-fast electron pump–probe experiments⁴⁶. A fully optical signature can provide direct evidence of bandgap renormalization: Fig. 4c shows a shift in the expected minimal energy of light absorption, appearing in addition to the Moss–Burstein shift that occurs in high doping⁴⁷. In addition, the same c.w. excitation can result in a steady-state emission spectrum that is both broadband and tunable, as shown in Fig. 4d. Such emission is a hallmark of the important role of transient time dynamics, because it cannot be explained by the quasi-thermal equilibrium that is usually reached in the conduction band when exciting at energies larger than the bandgap. To observe this emission, it should be coupled out from the near-field to the far-field, for example with a sharp metallic tip (coupling out of all momenta components, solid curves) or with a periodic grating (coupling out a narrow-momenta range, dashed curves).

The electron–polariton interaction is equivalent to an emitter that simultaneously interacts with a continuum of cavities, each having a different in-plane momentum. Each cavity obtains a different detuning energy and coupling constant and the system thus develops large decoherence, which dampens the nonlocal Rabi oscillations (that is, a form of T_2 relaxation). We quantify this effect by finding a correlation window function that expresses $C_{k_i}(t)$ directly as a function of its earlier times $C_{k_i}(t' < t)$, calculated analytically in Supplementary Section 2 and plotted in Supplementary Fig. 2. We show that maximal correlation times in our structure are of the order of 10 fs and are determined mostly by large changes in detuning energy between different plasmonic modes. Thus, we find that the plasmonic losses, their valence band absorption and external losses will not limit the bandgap renormalization until they reach a below 100 fs timescale. Extending this correlation time further could reduce the decoherence time in the system, which could open new applications for such systems in areas of quantum technologies. A promising avenue is to design the polaritonic dispersion curve to better match the band structure energies, which can be achieved, for example, with graphene nanoribbons. In such cases, the counter-rotating terms would cause a unique plasmonic statistics that could lead to novel nonlinear optics phenomena such as plasmonic blockade.

All previously known bandgap renormalization mechanisms differ from the one we present here in that they are based on intra-band interactions and are thus electrostatic in nature, or nearly so. Another renormalization mechanism that is analogous to our renormalization effect arises from the electron–phonon interaction³³ (for example, temperature-dependent bandgap renormalization). Other bandgap shifts can be caused by the environment, as a result of a potential barrier (as with thin spacers around QWs) and screening from nearby thin materials (for example, graphene^{48,49}) or many-body interactions in the material itself through various examples of electron–electron^{50,51} and electron–hole⁵² interactions. Some of these effects may be induced by ultra-fast laser pulses^{53–55}, which differ fundamentally from our vacuum-driven renormalization. Extending our work to such scenarios that involve many emitters, that is, many conduction electrons, may open up new effects of collective strong coupling and would require considering additional physics such as dark states and their coupling to the polaritons⁵⁶.

The physics of electron–polariton coupling in solids has exciting directions for future research and applications. On the practical side, the well-developed III–V layered fabrication processes, for example in FinFET transistors⁵⁷, may gain new degrees of freedom and tunability of their purely electronic properties by integration with nanophotonics and 2D polaritonic materials. Such transistors are especially exciting as a platform for tunable USC, with prospects for on-chip quantum technologies and for using the dominant polaritonic relaxation as a new passive optical heat dissipation mechanism.

On the fundamental side, our single-particle renormalization could be generalized to a density matrix approach to capture a density of excited electrons and thus allow us to explore new physics in out-of-equilibrium systems, as in refs. ^{53,58}. In addition, given that polaritons act as a mediator for the longer-range electron–electron interaction in solids, their USC with electrons might lead to new phenomena of long-range electron–electron scattering. Additionally, an intriguing yet admittedly speculative idea is to recall the phenomena of non-perturbative electron–phonon interactions, such as superconductivity, and attempt to recreate them through electron–polariton interactions in realistic nanophotonic structures that are designed to reach USC. Such phenomena could be externally induced on a wide range of materials by interfacing them with an optical environment that supports large-momentum polaritons. Ultimately, the simulation of phononic USC with polaritons could benefit from the polaritonic optical frequencies being above thermal excitations at room temperature, providing an avenue for collective quantum phenomena, such as superconductivity, at higher temperatures.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-020-0890-0>.

Received: 6 June 2019; Accepted: 31 March 2020;

Published online: 25 May 2020

References

1. Welton, T. A. Some observable effects of the quantum-mechanical fluctuations of the electromagnetic field. *Phys. Rev.* **74**, 1157–1167 (1948).
2. Coldren, L. A. & Corzine, S. W. *Diode Lasers and Photonic Integrated Circuits* (Wiley, 1995).
3. Sanchez-Mondragon, J. J., Narozhny, N. B. & Eberly, J. H. Theory of spontaneous-emission line shape in an ideal cavity. *Phys. Rev. Lett.* **51**, 550–553 (1983).

4. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
5. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
6. Haroche, S. & Kleppner, D. Cavity quantum electrodynamics. *Phys. Today* **42**, 24–30 (1989).
7. Khurgin, J. B. Excitonic radius in the cavity polariton in the regime of very strong coupling. *Solid State Commun.* **117**, 307–310 (2001).
8. Ciuti, C., Bastard, G. & Carusotto, I. Quantum vacuum properties of the intersubband cavity polariton field. *Phys. Rev. B* **72**, 115303 (2005).
9. Günter, G. et al. Sub-cycle switch-on of ultrastrong light–matter interaction. *Nature* **458**, 178–181 (2009).
10. Todorov, Y. & Sirtori, C. Intersubband polaritons in the electrical dipole gauge. *Phys. Rev. B* **85**, 045304 (2012).
11. Schwartz, T., Hutchison, J. A., Genet, C. & Ebbesen, T. W. Reversible switching of ultrastrong light–molecule coupling. *Phys. Rev. Lett.* **106**, 196405 (2011).
12. Scully, M. O. Collective Lamb shift in single photon Dicke superradiance. *Phys. Rev. Lett.* **102**, 143601 (2009).
13. Hümmer, T., García-Vidal, F. J., Martín-Moreno, L. & Zueco, D. Weak and strong coupling regimes in plasmonic-QED. *Phys. Rev. B* **87**, 115419 (2013).
14. Forn-Díaz, P., Lamata, L., Rico, E., Kono, J. & Solano, E. Ultrastrong coupling regimes of light–matter interaction. *Rev. Mod. Phys.* **91**, 025005 (2019).
15. Kockum, A. E., Miranowicz, A., De Liberato, S., Savasta, S. & Nori, F. Ultrastrong coupling between light and matter. *Nat. Rev. Phys.* **1**, 19–40 (2019).
16. Basov, D. N., Fogler, M. M. & García de Abajo, F. J. Polaritons in van der Waals materials. *Science* **354**, aag1992 (2016).
17. Low, T. et al. Polaritons in layered 2D materials. *Nat. Mater.* **16**, 182–194 (2016).
18. Rivera, N., Christensen, T. & Narang, P. Phonon polaritonics in two-dimensional materials. *Nano Lett.* **19**, 2653–2660 (2019).
19. Jablan, M., Buljan, H. & Soljačić, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
20. Giles, A. J. et al. Ultralow-loss polaritons in isotopically pure boron nitride. *Nat. Mater.* **17**, 134–139 (2017).
21. Ni, G. et al. Fundamental limits to graphene plasmonics. *Nature* **557**, 530–533 (2018).
22. Koppens, F. H. L., Chang, D. E. & García de Abajo, F. J. Graphene plasmonics: a platform for strong light–matter interactions. *Nano Lett.* **11**, 3370–3377 (2011).
23. Principi, A., van Loon, E., Polini, M. & Katsnelson, M. I. Confining graphene plasmons to the ultimate limit. *Phys. Rev. B* **98**, 035427 (2018).
24. Alcaraz Iranzo, D. et al. Probing the ultimate plasmon confinement limits with a van der Waals heterostructure. *Science* **360**, 291–295 (2018).
25. Rivera, N., Kaminer, I., Zhen, B., Joannopoulos, J. D. & Soljačić, M. Shrinking light to allow forbidden transitions on the atomic scale. *Science* **353**, 263–269 (2016).
26. Kurman, Y. et al. Control of semiconductor emitter frequency by increasing polariton momenta. *Nat. Photon.* **12**, 423–429 (2018).
27. Machado, F., Rivera, N., Buljan, H., Soljacic, M. & Kaminer, I. Shaping polaritons to reshape selection rules. *ACS Photonics* **5**, 3064–3072 (2018).
28. Lundeberg, M. et al. Tuning quantum non-local effects in graphene plasmonics. *Science* **357**, 187–191 (2017).
29. Christensen, T. et al. Nonlocal response of metallic nanospheres probed by light, electrons and atoms. *ACS Nano* **8**, 1745–1758 (2014).
30. Yan, W., Wubs, M. & Mortensen, N. A. Hyperbolic metamaterials: nonlocal response regularizes broadband supersingularity. *Phys. Rev. B* **86**, 205429 (2012).
31. Grosso, G. & Parravicini, G. P. *Solid State Physics* 2nd edn (Academic Press, 2013).
32. Giustino, F. Electron–phonon interactions from first principles. *Rev. Mod. Phys.* **89**, 015003 (2017).
33. Mahan, G. D. *Many-Particle Physics* 3rd edn (Springer, 2000).
34. Glauber, R. J. & Lewenstein, M. Quantum optics of dielectric media. *Phys. Rev. A* **43**, 467–491 (1991).
35. Scheel, S. & Buhmann, S. Y. Macroscopic QED—concepts and applications. *Acta Phys. Slov.* **58**, 675–809 (2008).
36. Vogel, W. & Welsch, D.-G. *Quantum Optics* 3rd edn (Wiley, 2006).
37. Sundaresan, N. M. et al. Beyond strong coupling in a multimode cavity. *Phys. Rev. X* **5**, 021035 (2015).
38. Polini, M. et al. Plasmons and the spectral function of graphene. *Phys. Rev. B* **77**, 081411 (2008).
39. Song, J. C. & Levitov, L. S. Energy flows in graphene: hot carrier dynamics and cooling. *J. Phys. Condens. Matter* **27**, 164201 (2015).
40. Rosker, M. J., Wise, F. W. & Tang, C. L. Femtosecond optical measurement of hot-carrier relaxation in GaAs, AlGaAs and GaAs/AlGaAs multiple quantum well structures. *Appl. Phys. Lett.* **49**, 1726–1728 (1986).
41. Jacob, Z., Smolyaninov, I. I. & Narimanov, E. E. Broadband Purcell effect: radiative decay engineering with metamaterials. *Appl. Phys. Lett.* **100**, 181105 (2012).
42. Poddubny, A., Iorsh, I., Belov, P. & Kivshar, Y. Hyperbolic metamaterials. *Nat. Photon.* **7**, 948–957 (2013).
43. Fei, Z. et al. Edge and surface plasmons in graphene nanoribbons. *Nano Lett.* **15**, 8271–8276 (2015).
44. Khurgin, J. B. Two-dimensional exciton–polariton—light guiding by transition metal dichalcogenide monolayers. *Optica* **2**, 740–742 (2015).
45. Khurgin, J. B. Hot carriers generated by plasmons: where are they generated and where do they go from there?. *Faraday Discuss.* **214**, 35–58 (2019).
46. Wang, K. et al. Coherent interaction between free electrons and cavity photons. Preprint at <https://arxiv.org/pdf/1908.06206.pdf> (2019).
47. Muñoz, M. et al. Burstein–Moss shift of n-doped In_{0.53}Ga_{0.47}As/InP. *Phys. Rev. B* **63**, 233302 (2001).
48. Raja, A. et al. Coulomb engineering of the bandgap and excitons in two-dimensional materials. *Nat. Commun.* **8**, 15251–15257 (2017).
49. Ugeda, M. M. et al. Giant bandgap renormalization and excitonic effects in a monolayer transition metal dichalcogenide semiconductor. *Nat. Mater.* **13**, 1091–1095 (2014).
50. Das Sarma, S., Jalabert, R. & Yang, S.-R. E. Band-gap renormalization in semiconductor quantum wells. *Phys. Rev. B* **41**, 8288–8294 (1990).
51. Walsh, A., Da Silva, J. L. & Wei, S. H. Origins of band-gap renormalization in degenerately doped semiconductors. *Phys. Rev. B* **78**, 075211 (2008).
52. Ryan, J. C. & Reinecke, T. L. Band-gap renormalization of optically excited semiconductor quantum wells. *Phys. Rev. B* **47**, 9615–9620 (1993).
53. Breusing, M., Ropers, C. & Elsaesser, T. Ultrafast carrier dynamics in graphite. *Phys. Rev. Lett.* **102**, 086809 (2009).
54. Pogna, E. A. A. et al. Photo-induced bandgap renormalization governs the ultrafast response of single-layer MoS₂. *ACS Nano* **10**, 1182–1188 (2016).
55. Ulstrup, S. et al. Ultrafast band structure control of a two-dimensional heterostructure. *ACS Nano* **10**, 6315–6322 (2016).
56. Gonzalez-Ballester, C., Feist, J., Badiá, E. G., Moreno, E. & Garcia-Vidal, F. J. Uncoupled dark states can inherit polaritonic properties. *Phys. Rev. Lett.* **117**, 156402 (2016).
57. Hisamoto, D. et al. FinFET—a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans. Electron Devices* **47**, 2320–2325 (2000).
58. Bostwick, A. et al. Observation of plasmarons in quasi-free-standing doped graphene. *Science* **328**, 999–1002 (2010).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

WW model using macroscopic QED. In this section, we use the WW model, as described in refs. ^{36,59}, to analyse the time-dependent light–matter interaction in a general case. Here, we include an arbitrary optical environment that can support multiple cavity modes or even a continuum of modes and a general quantum emitter that can have extended, spatially nonlocal, electronic wavefunctions. Specifically, we derive the evolution of a single electron in an excited state of the quantum emitter, which could decay to many possible final states while emitting a range of possible optical excitations. This new formalism assumes the existence of some range of final states that are unoccupied, an assumption that can be generalized by considering a Fermi–Dirac distribution at a finite temperature in a solid or an atom. We further assume that only one electron participates in the interaction—the electron that was initially in the excited state. In particular, we neglect the transitions of lower-energy electrons (more accurate at lower temperatures and in higher energy bands, where these electrons are further below). The system’s wavefunctions are fully described in equation (1) in the main text.

The Hamiltonian of the system is constructed in three parts: the interaction Hamiltonian in equation (2), the quantum emitter

$$H_{QE} = \sum_{\mathbf{f}} E_{\mathbf{f}} (\hat{b}_{\mathbf{f}}^{\dagger} \hat{b}_{\mathbf{f}}) + E_{\mathbf{i}} (\hat{b}_{\mathbf{i}}^{\dagger} \hat{b}_{\mathbf{i}}) \quad (7)$$

and the optical excitations

$$H_{EM} = \sum_{\mathbf{M}} \hbar \omega_{\mathbf{M}} \hat{a}_{\mathbf{M}} \quad (8)$$

The A^2 term is not considered in the interaction Hamiltonian, because its value is extremely small compared to that of the $\mathbf{p} \cdot \mathbf{A}$ term (Supplementary Section 6); however, this term, and additional extensions such as the RWA, can be added if our approach is used in other configurations. We stress that the analysis above can be generalized when adding more than a single initial excited state by considering higher-dimensional states and coupling constants to each possible combination.

The time evolution formulae are reached by solving the Schrödinger equation with Hamiltonian equations (7) and (8) for the state in equation (1). Differentiating equation (1) with respect to time, we obtain

$$\begin{aligned} \partial_t |\psi(t)\rangle &= (\hat{C}_{\mathbf{i}} - i\omega_{\mathbf{i}} \hat{C}_{\mathbf{i}}) e^{-i\omega_{\mathbf{i}} t} |\mathbf{i}, 0\rangle \\ &+ \sum_{\mathbf{f}} \sum_{\mathbf{M}} (\hat{C}_{\mathbf{f},\mathbf{M}} - i(\omega_{\mathbf{f}} + \omega_{\mathbf{M}}) \hat{C}_{\mathbf{f},\mathbf{M}}) e^{-i(\omega_{\mathbf{f}} + \omega_{\mathbf{M}}) t} |\mathbf{f}, 1_{\mathbf{M}}\rangle \end{aligned}$$

and by substituting this into the Schrödinger equation, $H|\psi(t)\rangle = i\hbar\partial_t|\psi(t)\rangle$, with the Hamiltonian in equations (7), (8) and (2). The two coupled differential equations for the amplitudes are then

$$\dot{\hat{C}}_{\mathbf{i}}(t) = -i \sum_{\mathbf{f}} \sum_{\mathbf{M}} e^{-i(\omega_{\mathbf{f}} + \omega_{\mathbf{M}} - \omega_{\mathbf{i}}) t} g_{\mathbf{i},\mathbf{f},\mathbf{M}} \hat{C}_{\mathbf{f},\mathbf{M}}(t) \quad (9)$$

$$\dot{\hat{C}}_{\mathbf{f},\mathbf{M}}(t) = -i e^{i(\omega_{\mathbf{f}} + \omega_{\mathbf{M}} - \omega_{\mathbf{i}}) t} g_{\mathbf{i},\mathbf{f},\mathbf{M}}^* \hat{C}_{\mathbf{i}}(t) \quad (10)$$

Finally, we integrate equation (10) over time with initial conditions $\hat{C}_{\mathbf{f},\mathbf{M}}(t=0) = 0$ and plug it into equation (9) to obtain an integro-differential equation for the initial excited-state amplitude, found in the first part of equation (4).

Up to this point, we have derived the WW time evolution integro-differential equations for an emitter with many final states that may emit into many possible optical excitations. It is at this point that we benefit from using the macroscopic QED description of the optical excitations; its advantage arises in calculating the expression (noting that $\tau = t - t'$)

$$\hbar^2 \sum_{\mathbf{M}} |g_{\mathbf{i},\mathbf{f},\mathbf{M}}|^2 e^{-i\omega_{\mathbf{M}} \tau} = \sum_{\mathbf{M}} \left| \langle \mathbf{i}, 0_{\mathbf{M}} | \frac{e}{m_0} \mathbf{p} \cdot \mathbf{A} | \mathbf{f}, 1_{\mathbf{M}} \rangle \right|^2 e^{-i\omega_{\mathbf{M}} \tau}$$

Importantly, the summation over the optical excitations $\sum_{\mathbf{M}}$ can also be written by integrating and summing over the triplet j, \mathbf{r}, ω , so that a single excitation creation is described by $|\mathbf{x}\omega j\rangle = \hat{f}_j^{\dagger}(\mathbf{x}, \omega)|0\rangle$ (refs. ^{35,60}), a notation that we use throughout this paper. For example, equation (8) can be written as $H_{EM} = \sum_{j=\mathbf{x},\mathbf{y},z} \int d\mathbf{r} \int d\omega \hbar \omega \hat{f}_j^{\dagger}(\mathbf{r}, \omega) \hat{f}_j(\mathbf{r}, \omega)$. Similarly, the identity operator can be written as $\sum_{\mathbf{M}} |1_{\mathbf{M}}\rangle \langle 1_{\mathbf{M}}| = \int d\omega \int d\mathbf{x} \sum_j |\mathbf{x}\omega j\rangle \langle \mathbf{x}\omega j|$, so that

$$\sum_{\mathbf{M}} \left| \langle \mathbf{i}, 0_{\mathbf{M}} | \frac{e}{m_0} \mathbf{p} \cdot \mathbf{A} | \mathbf{f}, 1_{\mathbf{M}} \rangle \right|^2 e^{-i\omega_{\mathbf{M}} \tau} = \frac{e^2}{m_0^2} \int d\omega \int d\mathbf{x} \sum_j |\langle \mathbf{i}, 0_{\mathbf{M}} | \mathbf{p} \cdot \mathbf{A} | \mathbf{f}, \mathbf{x}\omega j \rangle|^2 e^{-i\omega \tau}$$

To calculate the light–matter coupling constants, we use the macroscopic QED vector potential expression²⁵:

$$A_{\mathbf{k}}(\mathbf{r}) = \sqrt{\frac{\hbar}{\epsilon_0}} \int d\omega \frac{e}{c^2} \int d\mathbf{r}' \sqrt{\text{Im}\epsilon(\mathbf{r}', \omega)} \vec{\mathbf{G}}_{\mathbf{k}j}(\mathbf{r}, \mathbf{r}'; \omega) \hat{f}_j(\mathbf{r}', \omega) + \text{h.c.} \quad (11)$$

where $\vec{\mathbf{G}}$ is the classical dyadic Green’s function that expresses the entire optical environment and includes all possible optical excitations in the structure, and $\epsilon(\mathbf{r}', \omega)$ is the material’s permittivity at the excitation location. Using equation (11) and the commutation relations $[\hat{f}_j(\mathbf{r}, \omega), \hat{f}_k^{\dagger}(\mathbf{r}', \omega')] = \delta_{jk} \delta(\mathbf{r} - \mathbf{r}') \delta(\omega - \omega')$, we find

$$\begin{aligned} \sum_{\mathbf{M}} |g_{\mathbf{i},\mathbf{f},\mathbf{M}}|^2 e^{-i\omega_{\mathbf{M}} \tau} &= \frac{e^2}{\pi \epsilon_0 \hbar m_0^2} \int d\omega \int d\mathbf{x} \sum_j \frac{\omega^2}{c^4} \text{Im}\epsilon(\mathbf{x}, \omega_q) \\ &\langle \mathbf{i} | \vec{\mathbf{G}}_{\mathbf{k}j}(\mathbf{r}, \mathbf{x}; \omega) \mathbf{p}_{\mathbf{k}} | \mathbf{f} \rangle \langle \mathbf{f} | \vec{\mathbf{G}}_{\mathbf{l}j}(\mathbf{r}, \mathbf{x}; \omega) \mathbf{p}_{\mathbf{l}}^{\dagger} | \mathbf{i} \rangle e^{-i\omega \tau} \\ &= \frac{e^2}{\pi \epsilon_0 \hbar m_0^2} \int d\omega \int d\mathbf{x} \sum_j \frac{\omega^2}{c^4} \text{Im}\epsilon(\mathbf{x}, \omega_q) \int d\mathbf{r}' \psi_{\mathbf{i}}^*(\mathbf{r}') \vec{\mathbf{G}}_{\mathbf{k}j}(\mathbf{r}, \mathbf{x}; \omega) \\ &\quad \mathbf{p}_{\mathbf{k}} \psi_{\mathbf{f}}(\mathbf{r}') \int d\mathbf{r}'' \psi_{\mathbf{f}}^*(\mathbf{r}'') \vec{\mathbf{G}}_{\mathbf{l}j}(\mathbf{r}'', \mathbf{x}; \omega) \mathbf{p}_{\mathbf{l}}^{\dagger} \psi_{\mathbf{i}}(\mathbf{r}'') e^{-i\omega \tau}, \end{aligned}$$

where we have used Einstein’s summation convention and the electronic initial and final wavefunctions $\psi_{\mathbf{i},\mathbf{f}}(\mathbf{r})$. Next, we use two Green’s function identities, $\vec{\mathbf{G}}_{\mathbf{l}j}(\mathbf{r}', \mathbf{x}; \omega) = \vec{\mathbf{G}}_{\mathbf{j}l}(\mathbf{x}, \mathbf{r}'; \omega)$ and

$\frac{\omega^2}{c^2} \int d\mathbf{x} \text{Im}\epsilon(\mathbf{x}, \omega) (\vec{\mathbf{G}}(\mathbf{r}, \mathbf{x}, \omega) \vec{\mathbf{G}}(\mathbf{x}, \mathbf{r}', \omega))_{kl} = \text{Im} [\vec{\mathbf{G}}_{kl}(\mathbf{r}, \mathbf{r}', \omega)]$ (ref. ³⁵), to find the result shown in equation (3). This new result in equation (3) can be used in equation (4) to calculate the time evolution of the described energy state configurations, but is also valid more generally: our theory works for any optical environment and any electronic system, using the dyadic Green’s function and the electronic wavefunctions as the only inputs.

The dyadic Green function for graphene–semiconductor–metal structures.

The dyadic Green’s function, \mathbf{G} , has a cardinal role in determining the interaction between light and matter, as it contains the entire optical environment. The Green’s function is defined as the propagator of a dipolar excitation $\mathbf{p}(\mathbf{r}')$ into an electric field $\mathbf{E}(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = i\omega\mu_0 \int_V \vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{p}(\mathbf{r}') dV' \quad (12)$$

which satisfies the equation

$$\nabla \times \nabla \times \vec{\mathbf{G}} - \frac{\omega^2}{c^2} \vec{\epsilon}(\mathbf{r}, \omega) \vec{\mathbf{G}} = \delta(\mathbf{r} - \mathbf{r}') \mathbf{I} \quad (13)$$

Here, ω is the frequency, c is the speed of light, \mathbf{I} is the identity matrix, and $\vec{\epsilon}(\mathbf{r}, \omega)$ is the general electric permittivity tensor (which in this study we chose to be isotropic, $\epsilon \mathbf{I}$, but can be also anisotropic with the same formalism). In this Article, we exemplify our general formalism by considering an optical excitation from the semiconductor material layer in an air–graphene–semiconductor–perfect conductor heterostructure, as plotted in Supplementary Fig. 1a and in Fig. 2a. Because of the translational symmetry in the system, the dyadic Green’s function is expressed by a mode expansion according to the in-plane momentum $\mathbf{q} = (q_x, q_y)$ (refs. ^{61,62}):

$$\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{8\pi^2} \int dq \frac{1}{q^2(\hbar k_z)} [\mathbf{N}(\mathbf{q}, \mathbf{r}, \mathbf{r}') + \mathbf{M}(\mathbf{q}, \mathbf{r}, \mathbf{r}')] \quad (14)$$

where $\mathbf{N}(\mathbf{q}, \mathbf{r}, \mathbf{r}')$ and $\mathbf{M}(\mathbf{q}, \mathbf{r}, \mathbf{r}')$ are the spectral responses for the transverse electric (TE) and transverse magnetic (TM) polarizations, respectively, and $k_z = \sqrt{q^2 - \epsilon\omega^2/c^2}$ is the wavevector in the \hat{z} direction. In our problem, we consider only the response for the TM modes, as they include the large-momenta polaritons, which are dominant in the light–matter interaction compared to the small-momenta TE graphene plasmon modes⁶³.

In our system, there is no out-of-plane inversion symmetry that can further reduce our formulae. Thus, we need to consider four separate cases: where the dipolar excitation is located below the electric field location ($z > z'$), denoted by ‘+’, or above the electric field location ($z < z'$), denoted by ‘−’, and when the dipole is oriented along the in-plane or out-of-plane directions. Eventually, the TM spectral response achieves the form

$$\begin{aligned} \mathbf{M}(\mathbf{q}, \mathbf{r}, \mathbf{r}') &= M_{\parallel}(\mathbf{q}, \mathbf{r}, \mathbf{r}') \hat{\mathbf{q}} \hat{\mathbf{q}} + M_{\perp}(\mathbf{q}, \mathbf{r}, \mathbf{r}') \hat{\mathbf{z}} \hat{\mathbf{z}} \\ &= \frac{c^2 q^2 k_z^2}{\omega^2 \epsilon} e^{i\mathbf{q} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} F_{\perp}^{\parallel}(q, z, z') \hat{\mathbf{q}} \hat{\mathbf{q}} + \frac{c^2 q^4}{\omega^2 \epsilon} e^{i\mathbf{q} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')} \\ &\quad F_{\perp}^{\perp}(q, z, z') \hat{\mathbf{z}} \hat{\mathbf{z}} \end{aligned} \quad (15)$$

where the field and excitation in-plane locations are $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$. In cases where both z and z' are in the semiconductor material, the surrounding materials are seen only through the functions $F_{\perp}^{\parallel, \perp}(q, z, z')$, which include multiple reflections and transmissions (Supplementary Section 9):

$$F_{\perp}^{\parallel}(z, z') = 2 \sinh(k_z z') \cosh(k_z z) + 2 \sinh(k_z z') \sinh(k_z z) \frac{r_{\text{SE}} e^{-2k_z d} - 1}{r_{\text{SE}} e^{-2k_z d} + 1} \quad (16)$$

$$F_{\perp}^{\perp}(z, z') = 2 \sinh(k_z z) \cosh(k_z z') + 2 \sinh(k_z z') \sinh(k_z z) \frac{r_{\text{SE}} e^{-2k_z d} - 1}{r_{\text{SE}} e^{-2k_z d} + 1} \quad (17)$$

$$F_{\perp}^{\pm}(z, z') = -2 \cosh(k_z z') \sinh(k_z z) + 2 \cosh(k_z z') \cosh(k_z z) \frac{r_{\text{SE}} e^{-2k_z d} - 1}{r_{\text{SE}} e^{-2k_z d} + 1} \quad (18)$$

$$F_{\perp}^{\pm}(z, z') = -2 \cosh(k_z z) \sinh(k_z z') + 2 \cosh(k_z z') \cosh(k_z z) \frac{r_{gr} e^{-2k_z d} - 1}{r_{gr} e^{-2k_z d} + 1} \quad (19)$$

where $r_{gr} = \frac{\epsilon - 1 - \frac{\omega_{pl}^2}{\omega^2}}{\epsilon + 1 + \frac{\omega_{pl}^2}{\omega^2}}$ is the reflection coefficient from the semiconductor

to air via the graphene sheet, which is described by a nonlocal surface conductivity $\sigma_s(\omega, \mathbf{q})$ and includes plasmonic losses and Landau damping^{19,22,64}. Using the identity (proven in Supplementary Section 9)

$\text{Im}[\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)] \propto \int d\mathbf{q} e^{i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} \hat{\epsilon} \hat{\epsilon} \text{Im}[F(q, z, z')]$ for any polarization $\hat{\epsilon}$, we find

that the imaginary part of the dyadic Green's function in our structure for an in-plane oriented dipole becomes

$$\text{Im}[\vec{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)] = \frac{e^2}{4\pi\epsilon_0 a^2 \epsilon} \int d\mathbf{q} k_z \hat{\mathbf{q}} \hat{\mathbf{q}} e^{i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} \sinh(k_z z') \sinh(k_z z) \text{Im} \left[\frac{r_{gr} e^{-2k_z d} - 1}{r_{gr} e^{-2k_z d} + 1} \right] \quad (20)$$

Equation (20) describes the entire optical environment and can be substituted into equation (3) with any emitter (local or nonlocal) that is located in the semiconductor layer.

Data availability

Source data are available for this paper. All other data that support the plots in this paper and other findings of this study are available from the corresponding author.

Code availability

The code that supports the findings of this study is available from the corresponding author.

References

59. Weisskopf, V. & Wigner, E. Berechnung der natürlichen linienbreite auf grund der diracschen lichttheorie. *Z. Phys.* **63**, 54–73 (1930).

60. Knöll, L., Scheel, S. & Welsch, D.-G. in *Coherence and Statistics of Photons and Atoms* Ch. 7.3 (Wiley, 2001).
 61. Chew, W. *Waves and Fields in Inhomogeneous Media* (IEEE Press, 1995).
 62. Wasey, J. A. E. & Barnes, W. L. Efficiency of spontaneous emission from planar microcavities. *J. Mod. Opt.* **47**, 725–741 (2000).
 63. Lin, X. et al. Tailoring the energy distribution and loss of 2D plasmons. *N. J. Phys.* **18**, 105007 (2016).
 64. Gonçalves, P. A. D. & Peres, N. M. R. *An Introduction to Graphene Plasmonics* (World Scientific, 2016).

Acknowledgements

We thank D. Podolsky, N. Lindner and J.C. Song for their advice and for fruitful discussions regarding this paper. The research is supported by the Azrieli Faculty Fellowship, by the GIF Young Scientists' Program and by an ERC Starter Grant.

Author contributions

All authors made significant contributions to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41567-020-0890-0>.

Correspondence and requests for materials should be addressed to Y.K. or I.K.

Peer review information *Nature Physics* thanks Nicholas Christakis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.