DTU Library

Combining density functional theory with macroscopic QED for quantum light-matter interactions in 2D materials

Svendsen, Mark Kamper; Kurman, Yaniv; Schmidt, Peter; Koppens, Frank; Kaminer, Ido; Thygesen, Kristian S

Check for updates

Combining density functional theory with macroscopic QED for quantum light-matter interactions in 2D materials

Mark Kamper Svendsen [1,6✉], Yaniv Kurman [2,6], Peter Schmidt[3], Frank Koppens [3,4], Ido Kaminer [2] & Kristian S. Thygesen [5]

A quantitative and predictive theory of quantum light-matter interactions in ultra thin materials involves several fundamental challenges. Any realistic model must simultaneously account for the ultra-confined plasmonic modes and their quantization in the presence of losses, while describing the electronic states from first principles. Herein we develop such a framework by combining density functional theory (DFT) with macroscopic quantum electrodynamics, which we use to show Purcell enhancements reaching 10^7 for intersubband transitions in few-layer transition metal dichalcogenides sandwiched between graphene and a perfect conductor. The general validity of our methodology allows us to put several common approximation paradigms to quantitative test, namely the dipole-approximation, the use of 1D quantum well model wave functions, and the Fermi's Golden rule. The analysis shows that the choice of wave functions is of particular importance. Our work lays the foundation for practical ab initio-based quantum treatments of light-matter interactions in realistic nanostructured materials.

[1] CAMD, Department of Physics, Technical University of Denmark, Kgs. Lyngby, Denmark. [2] Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa, Israel. [3] ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona), Spain. [4] ICREA – Institució Catalana de Recerça i Estudis Avancats, Barcelona, Spain. [5] CAMD and Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, Kgs. Lyngby, Denmark. [6] These authors contributed equally: Mark Kamper Svendsen, Yaniv Kurman. ✉email: markas@dtu.dk

The development of van der Waals (vdW) heterostructures composed of atomically thin 2D layers has opened new opportunities for manipulating and enhancing light-matter interactions at the nanoscale. The very tight confinement of the electrons in these materials endows them with unique physical properties that can be altered and controlled by varying the number, type, and sequence of the atomic layers[1,2]. For example, the experiment, which served as motivation for the current theoretical work, demonstrated thickness-tunable infrared emission via intersubband transitions in vdW quantum wells (QWs) comprised of few-layer transition metal dichalcogenide (TMD) structures[3]. At the same time, vdW heterostructures can support light-matter hybrid modes such as plasmon-, phonon-, and exciton-polaritons[4,5] with ultra-confined electromagnetic (EM) fields in the direction normal to the 2D plane. The coupling between such polaritonic light modes and electron systems in nearby layers of the heterostructure provides an interesting avenue for controlling the optical properties of the material or even creating new types of quantum states—if sufficiently strong coupling can be realized. However, the theoretical description and, in particular, the practical computation, of such coupled quantum light-matter systems present significant challenges that call for new developments[6–8].

To set the stage for the following discussions, we briefly outline the most central notions and approximations relevant for the description of light-matter interactions in nanoscale materials, see Fig. 1. First, the light-matter interaction is often described within the dipole approximation, which is valid whenever the EM field can be assumed to be constant over the extent of the matter system, or more precisely, the extent of the relevant electronic wave functions. The dipole approximation works well in many cases but becomes problematic for confined EM modes, such as the plasmon polaritons considered in the present work[9,10]. In such cases, the detailed shape of the wave function (and the light field) must be taken into account when evaluating the light-matter interaction. The question of how detailed the description of the electron wave function must be remains largely unexplored, and obviously depends on the specifics of the considered system and the targeted accuracy. For the TMD/graphene heterostructures studied in this work, we find that the difference in the spontaneous emission rates obtained with model wave functions and full ab-initio wave functions, can differ by an order of magnitude and show clear qualitative differences as well.

When cavities or plasmonic structures are used to select and enhance distinct light modes, it is possible to reach the strong-coupling regime where the pure light and matter states become entangled. For this to happen, the light-matter coupling strength must exceed the damping rate of the light mode; or equivalently, the radiative lifetime of the matter state should be shorter than the lifetime of the light mode, such that emission-reabsorption processes can occur. We note in passing that the lifetime of the light mode is determined by the optical losses occurring in the relevant materials (see below). Under strong-coupling conditions it is necessary to diagonalize the light-matter Hamiltonian, or alternatively evolve the state dynamics, using non-perturbative methods (in practice within a truncated Hilbert space)[11–13]. However, in many cases of interest the coupling will be sufficiently weak that first-order perturbation theory, i.e. Fermi's Golden rule, applies. The combination of the dipole approximation and Fermi's Golden rule yields an electronic transition rate given by the product of the dipole strength and the local photonic density of states (DOS). In this case, owing to the equivalence of
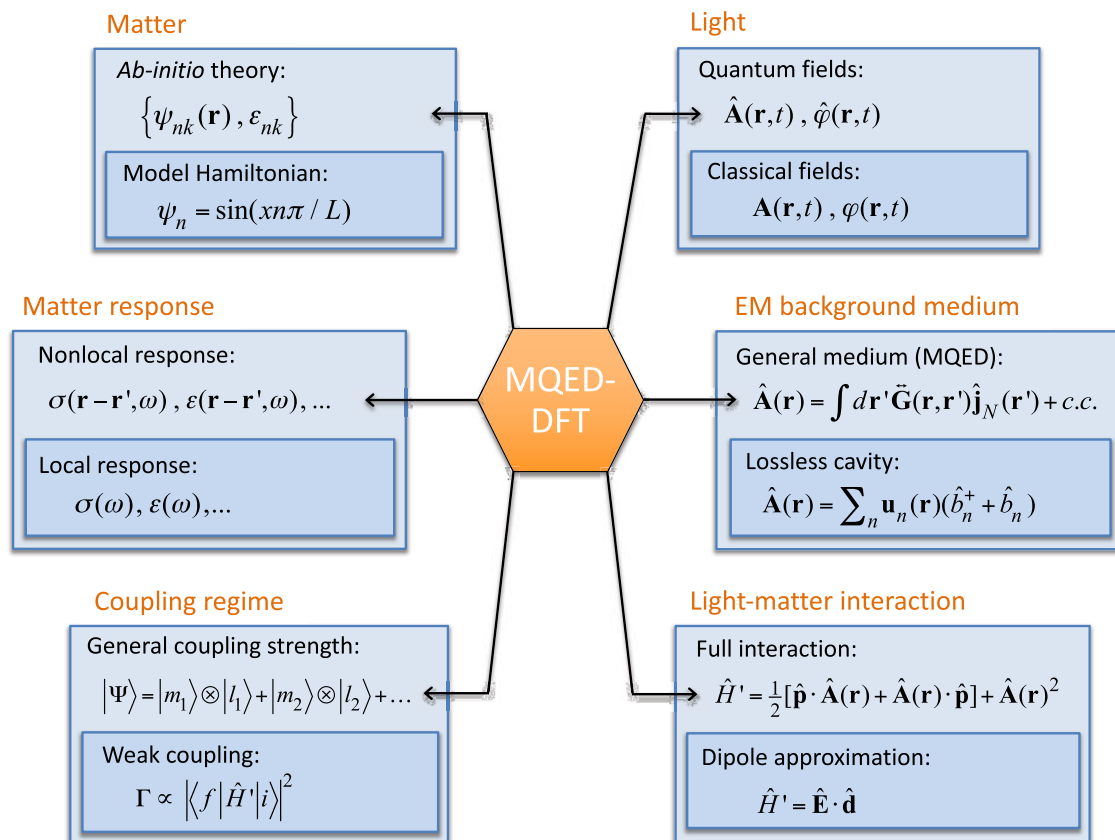

Fig. 1 Schematic illustration of the key quantities (blue boxes) entering the description of light-matter interacting systems. For each quantity, the most general/accurate level of the theory is shown at the top while commonly used approximations are shown below.
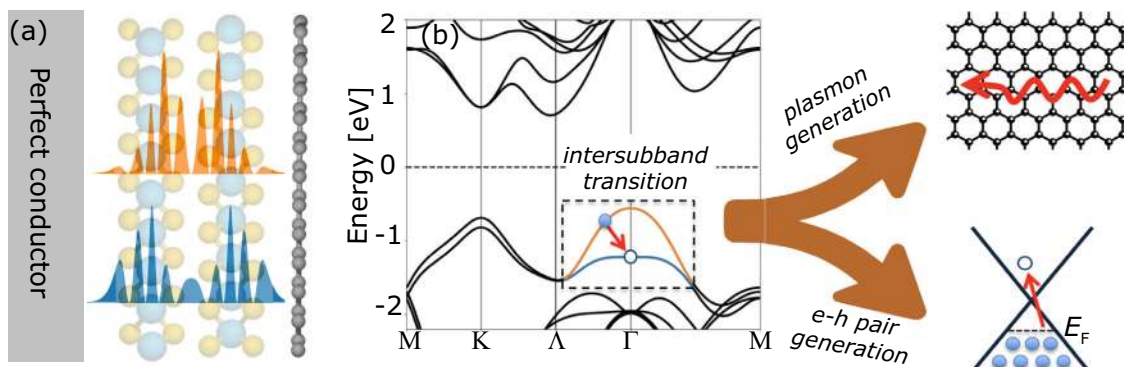
Fig. 2 Combining DFT and MQED in van der Waals materials. a Illustration of a 2-layer TMD quantum well sandwiched between a perfect conductor and a doped graphene sheet. The $z-$profile of two subband wave functions obtained from DFT are also shown. **b** The DFT band structure of bilayer WSe_2 with the two subbands at the Γ-point indicated in orange and blue. An electron from the filled upper subband can combine with a hole at the top of the lower subband by emitting either an electron–hole pair (loss) or a propagating acoustic plasmon-polariton.

the local photon DOS and the EM self-field of a point dipole, the spontaneous emission can be perceived as a semi-classical effect that depends on the electron wave functions only through the effective dipole strength. On the other hand, when the EM field varies on the scale of the electron wave function, the dipole-approximation breaks down[9], selection rules change[7,10] and the local semi-classical picture must be replaced by a full quantum treatment that explicitly incorporates the nonlocal nature of light-matter interactions.

The conventional approach to the quantization of the EM field is based on an expansion into orthogonal modes. However, in the presence of optical losses, i.e. when the conductivity of the EM background medium is finite in the relevant frequency range, the expansion into orthogonal modes is ill-defined and the quantization must be performed using alternative approaches such as macroscopic quantum electrodynamics (MQED)[14] or expansion in quasi-normal modes[15]. Finally, it is often necessary to invoke the optical response functions of the involved materials. This is for example the case for the MQED formalism to be discussed below. While spatially local, i.e. q-independent, response functions are often employed, and indeed work well in extended systems, this approximation becomes questionable in nanoscale structures such as 2D materials[16,17].

As will be shown, all of the common approximations used to describe quantum light-matter systems (see Fig. 1) can be violated in vdW QW devices. Consequently, these systems are ideal as testbeds for the testing and development of theoretical methods in this exciting field.

In recent years, there has been significant developments toward a full ab-initio description of coupled light-matter systems based on the fundamental Hamiltonian of non-relativistic QED, also known as the Pauli–Fierz Hamiltonian[18]. In particular, the quantum-electrodynamical density functional theory (QEDFT), which is a density functional type reformulation of the Pauli–Fierz Hamiltonian[19], has been introduced as a generalization of time-dependent density functional theory (DFT) to cases where the quantum nature of the light is important[18,20]. The fundamental nature of the QEDFT formalism has, for example, enabled studies of atomic emitters in cavities where the back-action of the EM field leads to "dressing" of the electronic states[21,22]. Although general, the QEDFT approach still has its limitations. An important one for making quantitative predictions of light-matter interactions in real materials, is the assumption of a lossless[23] or low-loss[24] EM background medium. Although this problem can in principle be alleviated by including the lossy environment as part of the active matter system[18] it is unlikely that this would be practical for complex nano-optical

setups, such as those involving vdW-heterostructures[1,6]. We further note that most applications of QEDFT have been restricted to the dipole approximation, although steps towards a more general description have recently been taken[24,25]. Finally, most QEDFT studies have been restricted to finite systems, such as atoms and small molecules, while the application to extended systems has been limited.

In this paper, we introduce an approach based on the Wigner–Weisskopf model to describe quantum light-matter interactions in complex 2D material setups under quite general conditions. Our methodology, combines MQED for quantizing the optical modes of an arbitrary optical environment with the single particle Kohn–Sham wave functions and energies from DFT for the electrons. This combination enables a quantitative description of the physics in a regime where quantized light modes, confined to a length scale of the electron wave function, interact strongly with the electrons while being subject to losses in the EM background medium. We use the method to calculate the radiative transition rates of a multilayer TMD QW sandwiched between graphene and a metal surface, see Fig. 2a. By varying the type of TMD material, the number of layers, and the graphene doping level, we determine the maximal Purcell enhancement achievable in such a setup to be 10^7, corresponding to intersubband transition rates of 4.5 THz with high efficiency of emission into the propagating plasmons. For the thinnest 2-layer stacks we observe somewhat lower Purcell factors of around 10^6, but higher absolute rates approaching 30 THz. These radiative lifetimes are at least an order of magnitude lower than the intrinsic (electron–phonon limited) lifetimes of electrons and holes in TMDs at low temperatures[3,26,27]. Consequently, the coupling to the graphene plasmons should indeed be the dominant mechanism governing the dynamics of the intersubband transitions at low temperatures. We compare our DFT-MQED results with common approximation schemes such as the dipole approximation and the use of 1D QW wave function models. We find quantitative differences in the obtained rates of up to an order of magnitude. In particular, we find that the rates are highly sensitive to the detailed spatial shape of the wave function, emphasizing the need for a proper description of the electronic states. Lastly, we explore cases where the first-order QED breaks down and higher-order processes must be considered to obtain a correct description of the Purcell effect.

Results

The MQED-DFT framework. The framework described in the following section is motivated by the new regimes of light-matter

interactions that are now accessible in nano-materials such as vdW heterostructures. Although the presented framework is general, we shall focus on the case of 2D layered geometries and apply the formalism to a TMD QW sandwiched between a conductor and a graphene sheet, see Fig. 2a. Such heterostructures feature rapid emission from the QW into acoustic graphene plasmons[28–30] that require a beyond-dipole MQED framework with ab-initio QW wave functions. We note that our work is the first to combine MQED with DFT. Moreover, the use of the Wigner–Weisskopf model enables us to analyze cases within the strong-coupling regime. A similar use of the Wigner–Weisskopf model can be found in[24,31], which also go beyond the Fermi's golden rule. We present here the main steps of our framework, while the details are described in Supplementary note 1.

We represent the light-matter interaction Hamiltonian in the Weyl gauge (which is defined by the requirement of a vanishing scalar potential),

$$\hat{H}_{\text{int}} = \frac{e}{m}\hat{\mathbf{A}} \cdot \hat{\mathbf{p}} - \frac{ie\hbar}{2m}\nabla \cdot \hat{\mathbf{A}} \ , \qquad (1)$$

where e is the electron charge, m is the electron mass, and $\hat{\mathbf{p}}$ represents the electron momentum operator. In writing Eq. (1) we neglected the diamagnetic $\hat{\mathbf{A}}^2$ term. The justification of this assumption and its implications are discussed in the "Limitations of MQED-DFT" subsection below. The Weyl gauge is an incomplete gauge and should therefore be accompanied by an additional condition to completely fix the gauge to avoid unphysical excitations arising in the quantization scheme[32]. We mention in passing that this condition is met because $\nabla \cdot (\overset{\leftrightarrow}{\epsilon} \mathbf{A}) = 0$ inside of the anisotropic TMD layer, the justification of which is discussed later. We further note that this means that in an isotropic dielectric material, our gauge reduces to the normal Coulomb gauge ($\nabla \cdot \mathbf{A} = 0$) common in quantum optics. In such cases, the $\nabla \cdot \hat{\mathbf{A}}$ term disappears from Eq. (1), leaving only the $\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}$ term. The $\nabla \cdot \hat{\mathbf{A}}$ term can thus be thought of as an anisotropic correction and we have generally found it to be small relative to the $\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}$ term.

Within MQED in the Weyl gauge, the vector potential is expressed as[33–35],

$$\hat{A}_i(\mathbf{r}) = \sqrt{\frac{\hbar}{\pi\epsilon_0}} \int d\omega \frac{\omega}{c^2} \int d^3s \ \overset{\leftrightarrow}{G}_{il}(\mathbf{r},\mathbf{s},\omega) \left[\sqrt{\text{Im } \epsilon(\mathbf{s},\omega)} \right]_{lj} \hat{f}_j(\mathbf{s},\omega) + \text{h.c.} \ , \qquad (2)$$

where $\hbar$ is the reduced Planck's constant, ϵ_0 is the vacuum permittivity, c is the speed of light, and the subscripts obey Einstein's summation convention. The quantization scheme is build around the Langevin noise current annihilation and creation operators, $\hat{J}_l(\mathbf{s},\omega) = \left[\sqrt{\text{Im } \epsilon(\mathbf{s},\omega)} \right]_{lj} \hat{f}_j(\mathbf{s},\omega)$ (and similar for $\hat{J}^\dagger$), with $\epsilon_{lj}(\mathbf{s},\omega)$ being the relative permittivity. The classical Dyadic Green's function (DGF), $\overset{\leftrightarrow}{G}(\mathbf{r},\mathbf{s},\omega)$, which satisfies

$$\left[\nabla \times \nabla \times - \left(\frac{\omega^2}{c^2} \right) \overset{\leftrightarrow}{\epsilon}(\mathbf{r},\omega) \right] \overset{\leftrightarrow}{G}(\mathbf{r},\mathbf{s},\omega) = \overset{\leftrightarrow}{\mathbf{I}} \ \delta(\mathbf{r}-\mathbf{s}), \qquad (3)$$

gives the electric field at position $\mathbf{r}$ generated by a point dipole of frequency ω at position $\mathbf{s}$ in the presence of an EM background medium with permitivity $\overset{\leftrightarrow}{\epsilon}$. In our case, the EM background permittivity is defined by the perfect conductor, the graphene sheet, and the TMD without contributions from the subband transitions (these define the active matter system, see below, and should be excluded from $\overset{\leftrightarrow}{\epsilon}$ to avoid double counting). In general, determining the DGF is a nontrivial task, which is often

accomplished using numerical methods[36]. However, in our case, the layered geometry of the vdW-heterostructure and the accompanying in-plane translation symmetry makes it possible to obtain a closed analytical solution for the DGF (see the Methods section).

The active matter system is defined by the electrons in the valence subbands near the Γ-point of the TMD QW, see Fig. 2b. We determine the electronic structure of the subbands of the isolated TMD structure by means of DFT calculations. DFT allows us to access the band structure and the single particle wave functions of the subbands[37], which in turn are used to evaluate the coupling matrix elements of the light-matter interaction Hamiltonian as described below. The details of the DFT calculation are provided in the Methods section.

Due to the in-plane translational invariance of the system, the electronic states of the subbands can be labeled by a subband index and a wave vector. In the following we consider the process where an electron from the upper subband (orange band in Fig. 2b) recombines with a hole at the top of the lower subband (blue) with the EM field initially in its ground state, i.e. a spontaneous emission process. The rate of this transition can thus be seen as the inverse lifetime of a hole in the lower subband due to the coupling to the EM field. Denoting the ground state of the TMD by $|\Psi_0\rangle$, the initial state representing a hole at the top of the lower subband (l) is written $|i\rangle = \hat{c}_{l,\mathbf{0}}|\Psi_0\rangle$, while the final state with a hole in the upper subband (u) at momentum $\mathbf{q}$ is given by $|f_\mathbf{q}\rangle = \hat{c}_{u,\mathbf{q}}|\Psi_0\rangle$.

The time-dependent combined state of the QW and EM field takes the form

$$\left|\Psi(t)\right\rangle = C_i(t)e^{-i\omega_i t}|i\rangle \otimes |0\rangle + \sum_\mathbf{q}\sum_{\nu_\mathbf{q}} C_{f,\nu_\mathbf{q}}(t)e^{-i(\omega_{\nu_\mathbf{q}}+\omega_{f_\mathbf{q}})t}\left|f_\mathbf{q}\right\rangle \otimes \left|\nu_\mathbf{q}\right\rangle \ , \qquad (4)$$

where $|\nu_\mathbf{q}\rangle$ labels a single photon state of the EM field, which due to the in-plane translational invariance of the layered structure considered in this work is described by a continuum of energies and in-plane momenta. We note that with the form of the many-body state in Eq. (4) we are limited to one-photon final states and thus we neglect multi-photon processes. Using the Wigner–Weisskopf approach, it can be shown that the time evolution of the initial state coefficient satisfies the equation of motion

$$\dot{C}_i(t) = -\int_0^t dt' \int d\omega \int d\mathbf{q} K(\omega,\mathbf{q})e^{-i(\omega-\omega_{if\mathbf{q}})(t-t')}C_i(t'), \qquad (5)$$

where $\hbar\omega_{if\mathbf{q}} = \varepsilon_{f\mathbf{q}} - \varepsilon_i$ is the electronic transition energy and the coupling kernel is given by

$$K(\omega,\mathbf{q}) = -\frac{1}{4\pi^2}\frac{e^2}{4m^2}\sum_{\nu_\mathbf{q}}\left|\left\langle f_\mathbf{q}\otimes\nu_\mathbf{q}|\hat{H}_{\text{int}}|i\otimes 0\right\rangle\right|^2\delta(\omega-\omega_{\nu_\mathbf{q}}). \qquad (6)$$

The Wigner–Weisskopf approach defined by Eqs. (5) and (6) describes the decay of the excited state into the EM environment and includes all optical one-photon decoherence channels. We note that via proper generalization other loss and decoherence sources, such as electron–phonon coupling, not considered in the current work, could in principle be accounted for by appropriate generalization of the many-body state and the coupling kernel. To account for pure dephasing a density matrix representation of the state is required and the Hilbert space would have to be extended to include the sets of states, $\{|i\rangle \otimes |\nu_\mathbf{q}\rangle\}$ and $\{|f_\mathbf{q}\rangle \otimes |0\rangle\}$.

The kernel, $K(\omega,\mathbf{q})$, in Eq. (6) describes the coupling strength between the electronic subband transitions and the optical excitations with frequency ω and in-plane momentum $\mathbf{q}$. To properly account for the nonlocality of the graphene, we employ the full nonlocal conductivity of the graphene[38] as described in
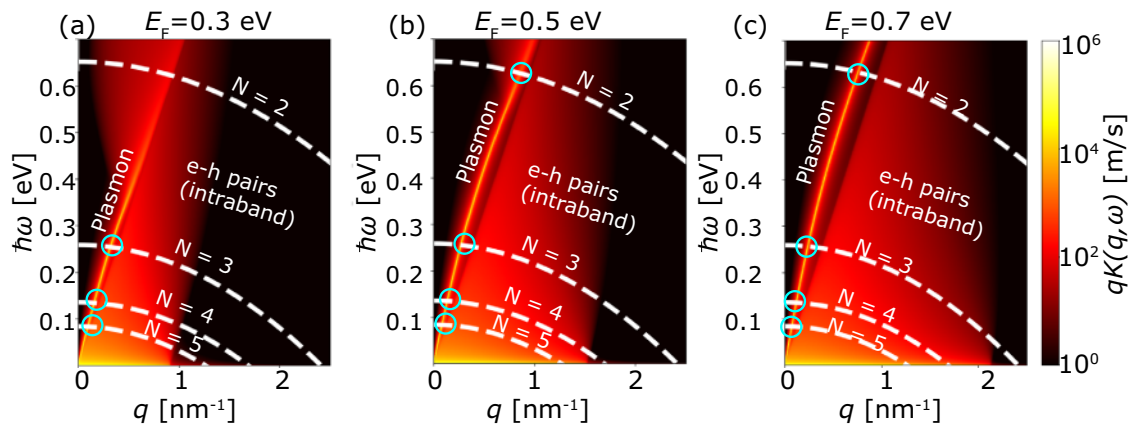
Fig. 3 Overlapping the electronic bandstructure and the plasmon dispersion to calculate their light-matter interactions. The coupling kernel for the transition between of holes between the lower and upper subband in a two-layer WSe$_2$ quantum well placed between a perfect conductor and a doped graphene sheet is shown for three different values of the graphene Fermi level, **a** 0.3 eV, **b** 0.5 eV and **c** 0.7 eV. The kernel describes the coupling into all optical excitations, including the plasmon and the electron–hole continuum as described by graphene's full nonlocal conductivity[38]. The dashed lines represent the non-vertical emission energies, $\hbar\omega_{if\mathbf{q}}$, derived from the DFT band structure for WSe$_2$ quantum wells of $N = 2, 3, 4,$ and 5 layers and the light blue circles denote the intersection between the dispersion of the intersubband transition energy and the graphene plasmon.

Methods. Importantly, we model the graphene via a nonlocal surface conductivity, neglecting the thickness of the graphene sheet. We note that treating the graphene as a 3D material with a finite thickness (with a bulk conductivity instead of a surface conductivity) should not affect the graphene plasmonic properties and thus not affect the results. The graphene conductivity depends strongly on the graphene Fermi level. The kernel therefore has to be calculated independently for each value of the graphene Fermi level. Exploiting the in-plane isotropy of the subband band structures, in Fig. 3 we show a color map of the coupling kernel integrated over the polar angle, $qK(\omega, q)$, for a 2-layer WSe$_2$ QW at three different values of the graphene Fermi level, alongside the transition energies, $\omega_{if\mathbf{q}}$, for the different numbers of layers. The acoustic graphene plasmon and the dissipative electron–hole (e–h) continuum are clearly visible.

From Eq. (6) it follows that the magnitude of the kernel at a given $(\omega, \mathbf{q})$ depends on two factors: the spatial overlap of the subband wave functions and the electric field of the optical mode (the matrix element), and the number of available optical modes at $(\omega, \mathbf{q})$ (the photonic DOS). The graphene plasmons affect the kernel via large matrix elements originating from their strong EM field. On the other hand, the e–h excitations in graphene individually produce weak EM fields but due to their large number they still can have a significant influence on the kernel, especially at small frequencies compared to the Fermi level.

There are several aspects of the formalism worth highlighting. First, we emphasize that the normal modes of the EM field, $|\nu_{\mathbf{q}}\rangle$, are used only symbolically in the above equations. In fact, as mentioned previously, the normal modes are not well defined in the presence of lossy media like the graphene sheet. Instead, one has to resort to so-called quasi-normal mode formalisms, of which MQED is one example, but others exist[39]. In practice, the MQED formalism allows us to obtain the kernel in terms of the DGF. Focusing on the dominant $\hat{\mathbf{A}} \cdot \hat{\mathbf{p}}$ term from the interaction Hamiltonian for simplicity, it can be shown that (see Supplementary note 1),

$$\sum_{\nu_{\mathbf{q}}} |\langle f_{\mathbf{q}} \otimes \nu_{\mathbf{q}} | \hat{H}_{\text{int}} | i \otimes 0 \rangle|^2 \delta(\omega - \omega_{\nu_{\mathbf{q}}}) = \iint d\mathbf{r}\, d\mathbf{r}'\, p_k^*(\mathbf{r}) \text{Im}\, \overset{\leftrightarrow}{G}_{kl}(\mathbf{r}, \mathbf{r}', \omega) p_l(\mathbf{r}'),$$

(7)

where we have defined the transition current $p_l(\mathbf{r}) = [\psi_{f\mathbf{q}}^*(\mathbf{r}) \hat{p}_l \psi_i(\mathbf{r})]$. By virtue of the DGF, an expansion in normal modes

becomes superfluous, and this resolves the problem of ill-defined normal modes. However, the replacement of the normal modes by the DGF is not only a technicality; it also leads to a somewhat different picture of the quantum light-matter interactions itself. The left hand side of Eq. (7) describes coupling of electronic states $|i\rangle$ and $|f_{\mathbf{q}}\rangle$ via the manifold of degenerate EM modes at frequency ω. In this picture, the EM field driving the electronic transition is simply the field of the EM vacuum, i.e., $\hat{\mathbf{A}}|0_{\text{EM}}\rangle$. In particular, it is independent of the state of the electrons. A quite different picture is suggested by the right-hand side. Recalling that $i\overset{\leftrightarrow}{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)$ is the electric field at $\mathbf{r}$ created by a point current source at $\mathbf{r}'$, we can interpret the double integral as the EM self-energy of the current distribution $\langle f_{\mathbf{q}} | \hat{\mathbf{p}}(\mathbf{r}') | i \rangle$ (we must use the imaginary part of the DGF to obtain the real part of the energy). We can further notice that $\langle f_{\mathbf{q}} | \hat{\mathbf{p}}(\mathbf{r}') | i \rangle$ is a measure of the quantum fluctuations of $\hat{\mathbf{p}}(\mathbf{r})$ in the initial and final states (the matrix element vanishes if one of the states is an eigenstate of $\hat{\mathbf{p}}(\mathbf{r}')$). Therefore, in this picture, the electronic transition is driven by an EM field whose source is the current density of the transition itself.

Limitations of MQED-DFT. Although the presented framework is quite general it still has its limitations. Most importantly, the framework is limited to the weak and strong coupling regimes, but not suited for treating cases of ultra-strong coupling. The first reason is that the framework is built around the Wigner–Weisskopf model which means that we are restricted to a one-photon Hilbert space and that we neglect the counter-rotating terms. Both of these assumptions are expected to be violated in the ultra-strong coupling regime[40]. Secondly, the fact that the wave functions and energies of the active matter system (the TMD subbands in the present case) are calculated independently of the vacuum fluctuations means that the states of the system might be misrepresented in situations where there is significant dressing of the electronic states by the vacuum field. In such cases, one should resort to a fully self-consistent method such as e.g., QEDFT. We stress, however, that such situations are only expected to arise in the case of ultra-strong coupling, which is very difficult to reach without external pumping of the EM field. Within the typical regimes of weak to strong coupling, in particular this includes systems in equilibrium, i.e., non-driven systems, we thus expect the MQED-DFT method to be valid and

provide an efficient alternative to more full-fledged ab-initio treatments.

To verify that the dressing of the subband states considered in this work is indeed small, we calculated the Lamb-shifts of the electronic energies and found that they never exceed 0.4% of the transition energy, (see Supplementary note 5). We thus conclude that the dressing of the electronic states by the EM vaccum field is neglible, and that a non-self consistent treatment of the light-matter interaction is justified for the systems studied in the present work.

Next we discuss the gauge fixing condition $\nabla \cdot (\overleftrightarrow{\epsilon} \mathbf{A}) = 0$. This condition is met since the MQED vector potential is defined in the absence of the intersubband transition and we assume that no other longitudinal sources exist inside of the TMD layer. The condition also follows directly from the form of the DGF presented in the methods section. This extra condition completely fixes the gauge and leaves us with a generalized version of the Coulomb gauge common in quantum optics. We note that the assumption that no other longitudinal sources exist is justified inside the TMD. At the TMD/graphene and TMD/conductor interfaces this condition is violated and the field develops a longitudinal component. However, the subband wave functions are well confined within the TMD region and have only negligible weights at the interfaces.

Next, we consider limitations arising from the chosen form of the interaction Hamiltonian used in this work. As mentioned in the above, in writing Eq. (1) we neglected the diamagnetic $\hat{\mathbf{A}}^2$ term. The diamagnetic current density, the $\hat{\mathbf{A}}^2$ term, has been neglected since it only describes two-photon emission/absorption and emission-reabsorption processes. The two-photon processes are not captured by our one-photon Hilbert space and neglecting them is consistent with the use of the Wigner–Weisskopf model. The emission-reabsorption processes leads to an energy renormalization of the states similar to the Lamb shift. We have calculated this shift using second-order perturbation theory for a few of the considered structures and found it to be below 0.4% of the undressed transition energies and thus negligible, see Supplementary note 4 for details on the calculation. While we believe that for the situation and method considered in this work the values calculated above give a good indication of the importance of the diamagnetic term, such an approach is only valid in a perturbative or finite level approach such as the restricted Hilbert space considered in this work[41]. More generally, the $\hat{\mathbf{A}}^2$ term becomes dominant when reaching the ultra-strong coupling regime[40] and any attempt at treating this regime would require an extension of the model to include the diamagnetic term. We further note that in any self-consistent treatment of both light and matter, e.g. QEDFT, the diamagnetic term is essential to ensure the stability of the coupled system even at weak coupling strengths[41] and any such approach must therefore naturally include the $\hat{\mathbf{A}}^2$ term.

Another limitation of our approach concerns the description of the EM field distribution via the DGF, which is based on a continuum model of the underlying materials, e.g. a (nonlocal) sheet conductivity for graphene and infinitely sharp interfaces between the materials. As a result, while the modes of the field resolve the mesoscopic structure of the vdW-heterostructure, the local fields due to the atomic structure are not included in the modes of the EM field.

Finally, we note that the assumption of a perfectly conducting metal substrate is an approximation. In practice, a finite conductivity would allow the field to penetrate into the metal and cause additional optical losses. While we expect these effects to be small in the spectral region considered here[29], they could be significant in e.g., the optical regime and can be included in the formalism at the price of a more elaborate (numerical rather than analytical) determination of the DGF.

Decay rates from MQED-DFT. In this section, we use the MQED-DFT framework to analyse the coupling of intersubband transitions to propagating graphene plasmons in the vdW QW devices depicted in Fig. 2. Throughout the discussion we shall distinguish the *radiative* decay rate due to emission of propagating plasmons from the *total* decay rate, which will also include the incoherent coupling to the e–h continuum in the graphene. The radiative rate is obtained by setting the coupling to the e–h continuum to zero (i.e. setting $K = 0$ everywhere in the (q, ω)-parameter space except in the region where the plasmon crosses the subband transition, see Fig. 3). Although the electron is emitting to different modes (for example different propagation directions), the interaction is expressed in terms of a single decay rate that incorporates all the emission channels.

In Fig. 4 we report the calculated radiative emission rates and Purcell factors for QWs of MoS$_2$, MoSe$_2$, WS$_2$, and WSe$_2$ as a function of the number of layers ($N = 2, ..., 5$) and the graphene doping level ($E_F = 0.1$–0.7 eV). In general, the four TMDs show very similar behavior with the (small) differences mainly originating from variations in the subband energy gaps. Figure 4a, b shows the maximum radiative emission rate and corresponding Purcell factor obtainable by tuning of E_F. The decay rates range from 2 to 28 THz, with the largest rates obtained for the thinnest structures. The corresponding lifetimes are at least an order of magnitude shorter than the intrinsic (electron–phonon limited) lifetimes of electrons and holes in TMDs at low temperatures[3,26,27]. Consequently, the coupling to the graphene plasmons should indeed be the dominant mechanism governing the dynamics of the intersubband transitions, at least in the low temperatures regime. The higher rates obtained for thinner structures is somewhat counter-intuitive as the thinner structures have smaller transition dipole moments. However, this effect is dwarfed by the larger number of plasmons available at higher energies (the 2D acoustic plasmon DOS scales as $\sim\omega$) where we find the transition energies of the thinner structures, see Fig. 3. Another way to think of this is that the plasmon is more confined at the larger transition frequencies of the thinner structures leading to stronger field enhancement[29].

While the absolute decay rates decrease with the number of layers, N, the Purcell factors show the opposite trend. This goes against the common Q/V-scaling rule for the Purcell enhancement and can be explained as follows: The (vertical) intersubband transition energy scales as $\omega_{ij}(N) \sim 1/N^2$. Within the point dipole approximation, the Purcell factor is essentially a measure of how much the photon DOS at the position of the emitter is enhanced relative to the free space value. While the 2D acoustic plasmon DOS scales as $\sim\omega$, the free space DOS scales as ω^2. This means that even though the plasmonic DOS increases with energy, its magnitude relative to free space DOS actually decreases with energy. This argument therefore explains why the Purcell enhancement decreases with decreasing thickness. Going beyond the dipole approximation, another argument for the increase in Purcell factor with N, is the change in the matrix element due to the spatial variation of the plasmon along the QW. As we will show in Fig. 5, variations from the dipole approximation do occur in the thicker structures. However, since the deviations from the dipole approximation are relatively small, we deduce that the dominant effect leading to the larger Purcell factor in the thicker structures is the relatively larger plasmonic DOS at the intersubband transition energy as compared to its photonic counterpart.
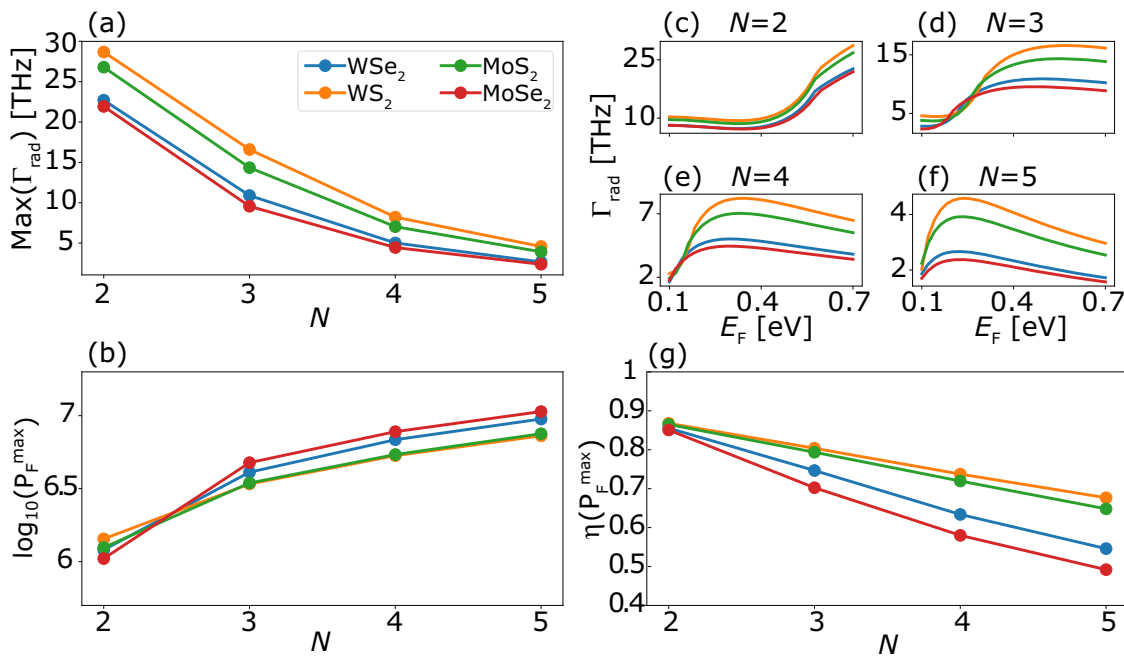
Fig. 4 The enhancement of transition rates as a function of number of layers and their composition. a The maximum radiative rates obtainable by adjusting the graphene Fermi level, for the different TMD materials and number of layers in the quantum well (N). **b** The Purcell factors corresponding to the maximal radiative rates in (**a**). Radiative emission rates as a function of graphene Fermi level for 2 (**c**), 3 (**d**), 4 (**e**), and 5 (**f**) layers. **g** Coupling efficiency, $\eta = \Gamma_{\mathrm{rad}}/\Gamma_{\mathrm{tot}}$, evaluated for E_F giving the maximum radiative rate. All rates were calculated using the generalized Weisskopf–Wigner MQED-DFT framework.
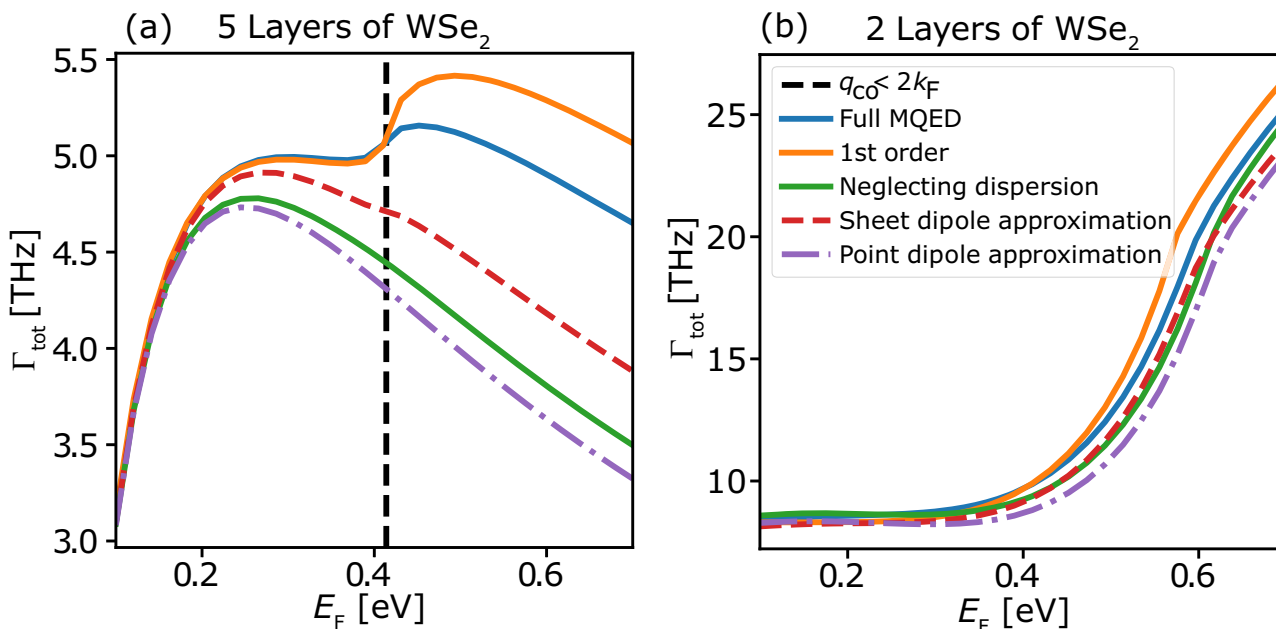

Fig. 5 Comparison of the total rates calculated using the five different approximations. The comparison is shown for 2 (**a**) and 5 (**b**) monolayers of WSe$_2$: The full time dependent MQED framework (blue), 1st order MQED perturbation theory (orange), the full MQED framework while neglecting the dispersion of the intersubband transition energy (green), approximating the TMD as a dipole sheet at $z_0 = d/2$ (red), and the point dipole approximation at $z_0 = d/2$ (purple). The dashed black line marks the graphene Fermi level at which horizontal transitions are opened up.

As can be seen from Fig. 4c–f, the rates show a strong dependence on the graphene doping level. This behavior is not surprising as the magnitude and dispersion of the plasmon EM field increases with the density of free carriers. This explains the initial rise of Γ_{rad} with E_F (except for the two-layer wells to which we return below). After the initial growth, the rates level-off and

even decrease as E_F is raised further. This happens because the dispersion curve of the graphene plasmon steepens as E_F is increased. This steepening of the plasmon dispersion at larger E_F results in a smaller transition momentum and thus smaller confinement of the plasmonic mode, leading to a reduction of the plasmonic DOS at the intersubband transition energy (blue circles

in Fig. 3). Within the range of E_F considered, the larger transition energies of the 2-layer stacks implies that the reduction of the plasmonic DOS due to the aforementioned dispersion steepening does not affect these thinnest structures, which explains why we do not observe a drop in the radiative rate for these systems.

Finally, we consider the radiative efficiency, $\eta = \Gamma_{rad}/\Gamma_{tot}$, evaluated for the graphene Fermi level giving the maximum Purcell enhancement, see in Fig. 4g. The efficiencies range from 0.5 to 0.9 with the highest values attained for the thinner structures when the Fermi level is chosen to maximize the radiative emission rate. This trend can again be traced to the higher subband transition energies in the thinner structures, which implies that the dispersion curve $\omega_{if}(q)$ avoids the e–h continuum, see Fig. 3. We note that since the metallic substrate is modeled as a perfect conductor there might be some additional losses arising from a real metallic substrate. However, we expect that these will be small in the infra-red region[29] and indeed we have found the corrections from a real model of the metal to be small relative to the rates found here. Therefore, to sum up this chapter, using our formalism we can deduce that the most effective and the fastest light-emitting TMD device should be formed by only a 2-layer TMD.

Assessment of common approximations. Having discussed the main results of the full MQED-DFT calculations, we now turn to a detailed analysis of the role played by the different elements of the method and a critical assessment of commonly used approximations. In this chapter we will focus on Fermi's golden rule, the dipole approximation, and the possible choices for modeling the electronic wave functions when going beyond the dipole approximation. The discussion will be centered around Fig. 5, which shows a comparison of the total emission rate calculated with various approximations for a 5-layer (a) and 2-layer (b) WSe$_2$ QW.

First-order perturbation theory. We first consider the applicability of first-order perturbation theory, i.e. Fermi's Golden rule, for calculating emission rates. In terms of the coupling kernel, the expression for the spontaneous emission rate takes the form

$$\Gamma = 2\pi \int_0^{2\pi} d\theta \int_0^{q_{hz}(\theta)} dq \; q \; K(\omega_{ifq}, q), \qquad (8)$$

which follows by making the Markov approximation in Eq. (5). In the momentum integral, the $q = 0$ limit corresponds to a vertical transition in the band structure while the limit $q = q_{hz}$ denotes the wave vector at which the transition becomes horizontal. The argument of the kernel expresses the conservation of energy and in-plane momentum in all transitions. Consequently, within the 1st order picture, the emission rate is thus assumed to be the sum of the kernel values along the arc (q, ω_{ifq}) (dashed white curves in Fig. 3).

The blue curve in Fig. 5a shows the reference result of the full MQED-DFT method while the orange curve shows the 1st order result from Eq. (8). As expected, the 1st order approximation works well for low E_F where the plasmon is weak and the density of e–h pairs is low. As E_F is raised, the light-matter coupling becomes stronger, and beyond a certain threshold, marked by the vertical dashed black line, there is a clear deviation from the full-coupling reference curve. In this case, the back-action of the all the emitted light modes on the TMD electron actually prohibits a faster emission and as a result the 1st order perturbation theory overestimates the rate. The un-perturbed decay rate is found from the exponential decay fit of the excited state probability decay trend (See Supplementary Figs. 4 and 5). The breakdown of the 1st order perturbation theory occurs as a result of the coupling to

low-energy horizontal e–h transitions across the graphene Fermi circle. Thus Fermi's golden rule is expected to fail when the horizontal QW transitions fall inside the e–h continuum, that is when $2k_F > q_{hz}$. For the 5-layer QW in Fig. 5a this criterion occurs at approximately $E_F = 0.4$ eV, while for the 2-layer structure in Fig. 5b it occurs outside the plotted scale.

Dipole approximations. The next approximation we investigate is the local dipole approximation. Within the dipole approximation, the graphene assisted EM vacuum field driving the emission process is assumed to vary little over the extent of the intersubband wave functions, and Eq. (7) reduces to

$$K(\omega, \mathbf{q}) \approx \hat{\mathbf{p}}_{if\mathbf{q}}^\dagger \cdot \text{Im} \, \overleftrightarrow{\mathbf{G}}(\mathbf{r}_0, \mathbf{r}_0, \omega) \cdot \hat{\mathbf{p}}_{if\mathbf{q}}, \qquad (9)$$

where $\mathbf{r}_0$ is a point in the center of the QW and $\hat{\mathbf{p}}_{if\mathbf{q}} = \langle f_\mathbf{q}|\hat{\mathbf{p}}|i\rangle$ is the dipole matrix element evaluated from the DFT wave functions. Equation (9) is the well known and widely used formula for spontaneous dipole emission[35]. Below we test this strictly local approximation against the fully nonlocal expression Eq. (7). We consider two different types of dipole approximations, namely the conventional *point-dipole* approximation and a *sheet-dipole* approximation. To model a point-dipole emitter we neglect the $\mathbf{q}$-dependence of the dipole matrix elements and transition frequencies and replace them on the right-hand-side of Eq. (9) by their values at the vertical ($q = 0$) transition. Unlike the point-dipole approximation, the sheet-dipole approximation takes the full $\mathbf{q}$-dependence of both dipole matrix elements (that is the QW wave functions) and transition frequencies into account, yet still evaluates the DGF at a single space point $\mathbf{r}_0$ in the center of the QW (See Supplementary S1 for more details). Thus, the sheet-dipole approximation becomes exact for EM fields with q-dependent energy dispersion and spatial variation along the in-plane direction but not in the normal direction.

The total decay rates obtained using the two dipole approximations are shown in Fig. 5 by the purple (point-dipole) and red (sheet-dipole) curves, respectively. Focusing on the 5-layer structure, we observe that both dipole approximations work well for small E_F. In this regime, the decay is dominated by the coupling to the plasmons whose electric field is long range and varies little over the extent of the QW, thus fulfilling the assumption of the dipole approximation. However, for larger values of E_F, the point-dipole approximation deviates from the full-coupling reference and underestimates the emission rate significantly. This difference arises mainly due to the neglect of the q-dispersion of the transition energies, which implies that the coupling to the low-energy e–h pairs is missed. Indeed, we obtain very similar results by setting $\omega_{if}(\mathbf{q}) = \omega_{if}(\mathbf{0})$ in the full MQED calculation (green curve). However, while the sheet dipole approximation indeed represents an improvement upon the point dipole approximation, it still falls short of the full MQED treatment. The q-dispersion of the transition energies alone is thus not enough to explain the shortcoming of the local treatment. Specifically, the coupling to the large q excitations in the e–h continuum of graphene is also underestimated in the sheet dipole approximation, due to the rapid spatial variation of their associated electric fields across the QW width, see Fig. 6d. In conclusion, any accurate description must therefore simultaneously account for the q-dispersion of the transition energies and describe the light-matter interaction beyond the dipole approximation. We note that both the neglect of q-dispersion and the dipole approximation have little influence on the rate obtained for the 2-layer QW, because the transition energy is much higher and the emission is dominated by the plasmon, see Fig. 3.
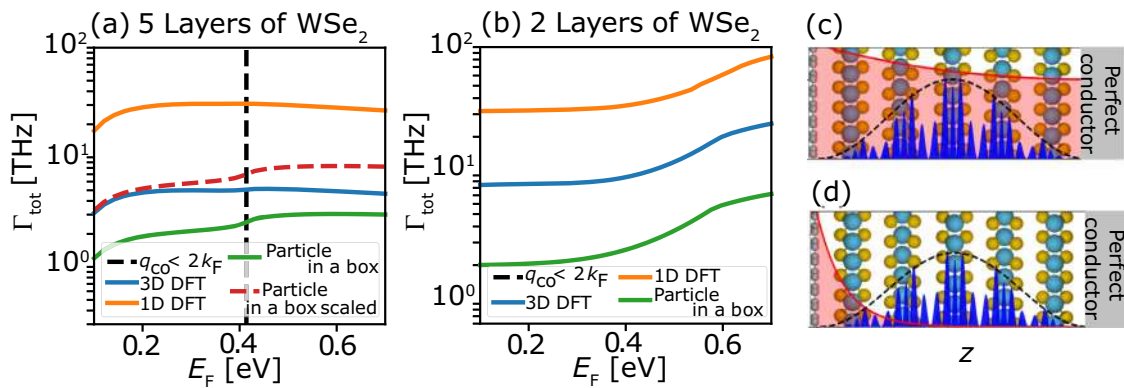
Fig. 6 Comparison of the total emission rates calculated using three different wave function approximations: particle in a box states, in-plane averaged DFT wave functions (1D DFT), and the full 3D DFT wave functions. a, **b** Comparison of the total emission rates calculated with the different wave function approximations for a 2-layer and 5-layer WSe$_2$ quantum well, respectively. Additionally, the dashed red line in (**a**) shows the rates calculated using the particle in a box wave functions scaled by the square of the ratio between the transition dipole moments calculated with the full 3D DFT wave functions and the particle in a box wave functions. The field enhancement profiles in red for the plasmon with in-plane wave vector $q = 0.2$ nm^{-1} (**c**) and $q = 1.21$ nm^{-1} (**d**). The latter corresponds to the wave vector for horizontal transitions, q_{hz}, of the 5-layer WSe$_2$ QW. The simple particle in a box wave function is shown by the dashed black line and the 1D DFT wave function is shown in blue.

Electronic wave functions. The final approximation that we investigate concerns the QW wave functions. As will be shown, different reasonable wave function choices can affect the rate by an order of magnitude. In Fig. 6a, b, we compare the total rates obtained using the full MQED framework with the full 3D DFT wave function (blue) to the results obtained using model wave functions of the form, $\psi(\mathbf{r}) = \chi(z)e^{i\mathbf{k}\cdot\mathbf{r}_\parallel}$, where $\chi(z)$ is either the in-plane averaged DFT wave function (orange) or a "particle in a box" approximation (green). To understand the role of the wave function, it is instructive to consider the explicit form of the light-matter coupling valid for our 2D geometry,

$$K(\mathbf{q},\omega) \approx C(\mathbf{q},\omega) \times \left| \int_{u.c.} d\mathbf{r}_\parallel \int dz u_{f_{\mathbf{q}}}(\mathbf{r}) \cosh\left(\sqrt{\frac{\epsilon_\parallel}{\epsilon_\perp}}qz\right) \partial_z u_i(\mathbf{r}) \right|^2 . \quad (10)$$

In this expression u denotes the (in-plane) periodic part of the QW wave functions and $C(\mathbf{q}, \omega)$ is independent of the wave functions (it contains the density and field strength of the graphene assisted EM vacuum modes, see the Supplementary note 1).

The different wave function approximations produce two types of errors. The first type of error is a simple error in the effective dipole. This error occurs when the q of all of the relevant excitations fulfill the condition $q \ll 1/d$ (d is the QW width) such that the field is approximately constant over the width of the QW. This condition is fulfilled for the 2-layer well in Fig. 6a. In this case, the different wave functions produce different effective dipoles but with no qualitative effect on the rate as function of E_F. As shown in the Supplementary note 2, the three curves in Fig. 6a are to a very good approximation related by a simple rescaling by the square of their respective transition dipole moments.

The second type or error occurs when the wave functions are coupling to large q excitations. In this case, the coupling becomes sensitive to the spatial shape of the wave functions. This is for example the case for the 5-layer structure, where the rate from the particle-in-a-box wave function deviates qualitatively from the DFT results for higher values of E_F. This can be clearly seen if we compare the rates calculated using the full 3D DFT wave functions to the rates calculated using the particle in a box states rescaled by the ratio of the effective dipole moments, blue and dashed red lines in Fig. 6a respectively. Specifically, the qualitative deviation happens when $2k_F$ becomes comparable to q_{hz}, as

indicated by the vertical dashed line. At this point, strong quenching due to the coupling of the horizontal transition to the low-energy e–h pairs becomes possible. Importantly, however, the strong quenching is not observed for the DFT wave functions because they are localized further inside the TMD structure (further away from the interface with the graphene) as compared to the standing waves, and therefore their overlap with the vacuum field associated with the $2k_F$ e–h pairs is significantly suppressed, see Fig. 6c.

Discussion

In this paper, we have combined MQED with DFT to obtain a computationally efficient framework for quantitatively describing quantum light-matter interactions in solid-state systems. The active quantum matter is represented by its DFT band structure and wave functions while the EM background media are described by their appropriate macroscopic linear response functions, which can be spatially nonlocal and dissipative. While the Kohn–Sham energies and wave functions from DFT are expected to provide a good approximation to the true subband states, it should be mentioned that they are subject to the limitations of the specific exchange-correlation functional employed and the Kohn–Sham scheme in general. In the future it would thus be interesting to explore how the use of other wave function schemes, for example hybrid functionals or GW self-energies, would affect emission rates and other quantum light-matter properties. The approach taken in the present work is in many ways complementary to the QEDFT formalism. While being less fundamental in nature than the QEDFT, our method avoids many of the approximations commonly made in the description of coupled light-matter systems, c.f. Fig. 1, while it remains applicable to relatively large/complex systems. Instead of taking the ab-initio route directly, we take advantage of the fact that the quantization of the light degrees of freedom in interacting, potentially lossy light-matter systems can be achieved within the MQED formalism[14,33]. A powerful feature of the MQED framework is that it treats all light modes on an equal footing. Specifically, it achieves the simultaneous quantization of lossy—as well as undamped propagating modes. While the coupling to propagating modes is almost always the main objective, in practice the optical losses are often the performance-limiting factor. Since losses are present and often of significant practical importance in real materials, they are essential to include in a

quantitative theory. Even within the MQED formalism, and similar quantization schemes of lossy quasi-normal modes[15], the light-matter interaction is usually described through the dipole approximation, i.e. the relevant electronic transition matrix elements are replaced by point dipole emitters[14]. However, the vdW heterostructures studied in the present work, comprise a clear example where the dipole approximation breaks down when emitting lossy optical modes. Furthermore, as shown in this work, the quantum description is essential to capture the deviations from first-order perturbation theory, which we observe when the carrier concentration in the plasmon-carrying graphene sheet becomes sufficiently high.

We have applied the combination of MQED and DFT to study the coupling of intersubband transitions in atomically thin TMD QWs to the optical modes of a doped graphene sheet. Our calculations predict radiative lifetimes below 1 ps and Purcell factors approaching 10^7. Importantly, these lifetimes are an order of magnitude shorter than the intrinsic low temperature phonon-limited carrier lifetimes in TMDs and thus the spontaneous emission of graphene plasmons from intersubband transitions should indeed be efficient and observable. We note that Purcell factors as high as the ones reported here are unusual for extended 2D structures, and were previously only reported for geometries confined in all three dimensions[42–44]. In contrast to the case of 3D confined cavities where the enhanced emission is due to the coupling to a distinct EM cavity mode, the strong enhancement of the 2D subband transitions considered here is enabled by the significant spatial overlap of the subband wave functions with the highly confined acoustic graphene plasmons.

Although the calculated Purcell factors are indeed very high, they are at least two orders of magnitude smaller than the conventional Q/V enhancement rule highlighting the limitations of the latter. The deviations from the Q/V rule are in part connected to the failure of both the first-order perturbation theory, and the dipole approximation in the examined structures. Our formalism allowed us to benchmark these approximations, which fail in our structure when coupling to high-momentum lossy e–h pair excitations in the graphene that increase the total coupling and produce a significant field variation across the TMD QW. Other common approximations, such as simpler wave function models can lead to decay rates differing by more than a factor of 10 depending on the details of the model. Apart from the difference in dipole moments, the finer details in the spatial shape of the wave function can affect the coupling to the high-q lossy modes in graphene resulting in a wrong qualitative behavior of e.g. the rate versus graphene doping level.

The developed MQED-DFT method is computationally compact, methodologically simple, and is not limited to a specific system geometry or material type. In addition, the MQED quantization scheme can deal with situations where the emitter is interacting with lossy media, e.g. plasmonic metal structures, doped semiconductors, or insulators with polar phonons, which can be included in macroscopic optical parameters of the EM background media. On the other hand, the non-self consistent treatment of the light-matter interaction limits the method to the weak to strong coupling regimes where the dressing of the electronic states by the EM field can be neglected. In the ultra-strong coupling regime, the EM field dresses the electronic states, and one must resort to fully self-consistent treatments such as QEDFT. Alternatively, our current theory could be extended using the Power–Zienau–Woolley gauge that overcome several of our approximations when using an electronic multi-polar term[45]. However, in the most typical cases of weak-strong coupling, the method developed here provides an efficient and practical alternative to methods like QEDFT, which is applicable to realistic setups in quantum nano-photonics.

Methods

DFT calculations. The wave functions of the TMD stacks used to evaluate the light-matter coupling kernel Eq. (5) were obtained from DFT employing the PBE exchange-correlation (xc) functional[46] and a plane wave cut-off of 800 eV. To resolve the subband structure a Γ-point centered 60×60 k-point grid was employed. To avoid spurious interaction between periodically repeated supercells, a 20 Å vacuum was included perpendicular to the TMD stack. The relaxed atomic structures of the four TMD monolayers were obtained from the C2DB database[47]. Next, N-copies of the monolayer were stacked in the 2H-configuration and the interlayer distance was relaxed using the C09-vdW xc-functional[48] while keeping the intralayer structure fixed. All calculations were performed with the GPAW code[49]. The graphene sheet and the perfect mirror were both placed at a distance of $(d_{Gra} + d_{TMD})/2$ from the outer most atomic layer of the TMD stack, with d_{Gra} and d_{TMD} being the interlayer distances in graphene and the TMD respectively.

1D wave function approximations. In addition to the full 3D DFT generated wave functions, this work considers two different wave function approximations: Particle in a box states, and the in-plane average of the DFT states. The particle in a box states are given as,

$$\phi_n(\mathbf{r}) = \left(\tfrac{2}{d}\right)^{1/2} \sin\left(\tfrac{n\pi z}{d}\right) , \tag{11}$$

where n is the subband index and d is the width of the TMD layer. The 1D DFT states are found when approximating $\phi_n(\mathbf{r})$ with the in-plane average of the full 3D wave functions.

The Dyadic Green's function. In this section we will find the Dyadic Green's function(DGF) for the structure discussed in the main text. The DGF, $\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)$, provides a general description of the EM environment of a structure, and it plays a fundamental role in determining the interaction between matter and the modes of the EM field. $\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)$ is defined as propagating a current density, $\mathbf{j}(\mathbf{r}', \omega)$, at a source points $\mathbf{r}'$, onto the corresponding electric field at point $\mathbf{r}$ according to:

$$\mathbf{E}(\mathbf{r}) = i\omega\mu_0\mu(\mathbf{r}) \int_V \overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}')\mathbf{j}(\mathbf{r}')dV' \tag{12}$$

This translates into the DGF itself being defined as the electric field $\mathbf{E}$ at point $\mathbf{r}$, generated by a radiating dipole, $\mathbf{p}$, located at the source point $\mathbf{r}'$. Therefore, the field from the radiating dipole can be expressed as, $\mathbf{E}(\mathbf{r}, \omega) = \omega^2\mu_0\mu \overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega) \cdot \mathbf{p}$[35]. Specifying to non-magnetic materials, $\mu = 1$, the equation of motion for the DGF is the Helmholtz equation with a point source term:

$$\left[\nabla \times \nabla \times - \left(\frac{\omega^2}{c^2}\right)\overleftrightarrow{\epsilon}(\mathbf{r}, \omega)\right]\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r} - \mathbf{r}') \tag{13}$$

where ω is the angular frequency and c is the speed of light. In this work, we consider the electric permittivity tensor, $\overleftrightarrow{\epsilon} = \text{diag}(\epsilon_\parallel, \epsilon_\parallel, \epsilon_\perp)$ to account for anisotropic dielectric properties of the TMD stacks.

Because of the layered nature of the geometry, the in-plane translational invariance of the system enables the DGF to be expressed as a mode expansion over the in-plane momentum, $\mathbf{q} = (q_x, q_y)$ in the following way[35,50,51]:

$$\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega) = \int d\mathbf{q} e^{i\mathbf{q}\cdot(\rho - \rho')}[\overleftrightarrow{\mathbf{N}}(\mathbf{q}, z, z') + \overleftrightarrow{\mathbf{M}}(\mathbf{q}, z, z')] \tag{14}$$

where $\overleftrightarrow{\mathbf{N}}(\mathbf{q}, z, z')$ and $\overleftrightarrow{\mathbf{M}}(\mathbf{q}, z, z')$ describes the spectral response of the TE and TM polarizations respectively. However, since the cavity plasmon mode is TM polarized, we retain only the TM polarized part of the DGF.

To further derive the explicit form of the Green's function in the multilayered structure, we use express $\overleftrightarrow{\mathbf{M}}(\mathbf{q}, z, z')$ with the reflection coefficients from each side of the TMD[50,51]. It can be shown that the components of the TM response are

$$\text{Im}\left[M_{\hat{q}\hat{q}}(\mathbf{q}, z, z')\right] = \frac{k_z c^2}{4\pi^2\omega^2\epsilon_\parallel}\sinh(k_z z)\sinh(k_z z')\text{Im}\left[\frac{r_g e^{-2k_z d} - 1}{r_g e^{-2k_z d} + 1}\right] \tag{15}$$

$$\text{Im}\left[M_{\hat{z}\hat{q}}(\mathbf{q}, z, z')\right] = \frac{iqc^2}{4\pi^2\omega^2\epsilon_\perp}\cosh(k_z z)\sinh(k_z z')\text{Im}\left[\frac{r_g e^{-2k_z d} - 1}{r_g e^{-2k_z d} + 1}\right] \tag{16}$$

$$\text{Im}\left[M_{\hat{q}\hat{z}}(\mathbf{q}, z, z')\right] = \frac{-iqc^2}{4\pi^2\omega^2\epsilon_\perp}\sinh(k_z z)\cosh(k_z z')\text{Im}\left[\frac{r_g e^{-2k_z d} - 1}{r_g e^{-2k_z d} + 1}\right] \tag{17}$$

$$\text{Im}\left[M_{\hat{z}\hat{z}}(\mathbf{q}, z, z')\right] = \frac{q^2 c^2\epsilon_\parallel}{4\pi^2\omega^2 k_z\epsilon_\perp^2}\cosh(k_z z)\cosh(k_z z')\text{Im}\left[\frac{r_g e^{-2k_z d} - 1}{r_g e^{-2k_z d} + 1}\right] \tag{18}$$

where $k_z = \sqrt{\frac{\epsilon_\parallel}{\epsilon_\perp}q^2 - \epsilon_\parallel\frac{\omega^2}{c^2}}$ is the wave vector in the out of plane direction, and $\hat{q}$ is an in plane spatial unit vector pointing in the direction of $\mathbf{q}$. In addition, we have assumed perfect reflection from the TMD to the metal and used the quasistatic reflection coefficient $r_g = \frac{\sqrt{\epsilon_\parallel\epsilon_\perp} - 1 - i\frac{q\sigma_s(q,\omega)}{\epsilon_0\omega}}{\sqrt{\epsilon_\parallel\epsilon_\perp} + 1 + i\frac{q\sigma_s(q,\omega)}{\epsilon_0\omega}}$ as the TMD to Graphene reflection coefficient (see

Supplementary note 3 for derivation) with $\sigma_s(q, \omega)$ as the nonlocal graphene surface conductivity which includes plasmonic losses and Landau damping[38]. We note that the Green's function expressions can be derived directly from the electric fields from a dipole oriented in the $\hat{\mathbf{q}}$ and $\hat{\mathbf{z}}$ directions. Finally, for our further calculations we will also use the identity, $\mathrm{Im}[\overleftrightarrow{\mathbf{G}}(\mathbf{r}, \mathbf{r}', \omega)] = \int d\mathbf{q} e^{i\mathbf{q}\cdot(\rho-\rho')}\mathrm{Im}[\overleftrightarrow{\mathbf{M}}(\mathbf{q}, z, z')])]$, which describes the entire TM optical response that the TMD experiences, whether we want to consider the TMD emitter as local, nonlocal or even a dipole emitter.

Background permittivity

TMD permittivity. The permittivities of the TMD's were calculated both in-plane and out-of-plane using the GPAW code[49] and they are consistent with the values found in the literature[52]. The permittivities are chosen not to include the inter-subband transitions as to avoid double counting and are thus frequency independent. The values used are shown in Table 1.

Nonlocal graphene conductivity. In order for the light-matter interactions to take into account all possible optical excitations in the graphene, including the high-momentum e–h continuum, there is a necessity for a nonlocal graphene conductivity as opposed to a simple Drude model. This approach include any possible optical response in the graphene, within the random phase approximation. In this chapter, we will explain how to retrieve the nonlocal graphene surface condictivity in a similar manner as found in Peres and Goncalves[38]. We note that treating the graphene as a 3D material with a finite thickness (with a bulk conductivity instead of a surface conductivity) should not affect the graphene plasmonic properties. First, the graphene conductivity, $\sigma(q, \omega)$, is defined in the following way:

$$\sigma(q, \omega) = \frac{4i\sigma_0 \hbar \omega}{q^2} \chi_t\left(\frac{q}{k_F}, \frac{\hbar\omega}{E_F}\right) \quad (19)$$

with $\sigma_0 = \frac{e^2}{4\hbar}$, and χ_t being the susceptibility of normalized variables, defined as:

$$\chi_t(x, y) = \frac{\left(1 + i\frac{\Gamma}{yE_F}\right)\chi_g\left(x, y + i\frac{\Gamma}{E_F}\right)}{1 + i\frac{\Gamma}{yE_F}\chi_g\left(x, y + i\frac{\Gamma}{E_F}\right)/\chi_g(x, 0)} \quad (20)$$

with Γ as the graphene dampening parameter. This work uses the value of Γ to be 16 meV which is consistent with choice by Dias et al.[53]. Finally $\chi_g(x, y)$ is defined according to the regions in Fig. 7, that describe the regions of different optical responses in the graphene. For more readable expressions, we define the following auxiliary functions, $F(\xi) = \xi\sqrt{\xi^2 - 1} - \mathrm{arc\,cos\,h}(\xi)$, and, $C(\xi) = x\sqrt{1 - \xi^2} - \mathrm{arc\,cos}(\xi)$. Furthermore, $x = q/k_F$ is the normalized momentum and $y = \hbar\omega/E_F$ is the normalized energy.

The different regions of the responses are:

Region 1A. Region 1A is defined as the part of the (q, ω) plane where $x > \mathrm{Re}[y]$ and $\mathrm{Re}[y] + x < 2$. In this region the graphene electrons stay in the same band while changing largely their momentum. The expression for the nonlocal susceptibility is:

$$\chi_g^{1A}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} + i\frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{x^2 - y^2}}\left[F\left(\frac{2-y}{x}\right) - F\left(\frac{2+y}{x}\right)\right] \quad (21)$$

Region 2A. Region 2A is defined by $x - \mathrm{Re}[y] < 2$, $x > \mathrm{Re}[y]$ and $x + \mathrm{Re}[y] > 2$. In this region the graphene electrons change their band while also changing largely their momentum. The expression for the susceptibility is:

$$\chi_g^{2A}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} - \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{x^2 - y^2}}C\left(\frac{y-2}{x}\right)$$
$$- i\frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{x^2 - y^2}}F\left(\frac{2+y}{x}\right) \quad (22)$$

Region 3A. Region 3A is defined by $x - 2 > \mathrm{Re}[y]$. In this region there is no direct transition for the graphene and thus the optical response is weak. The expression for the susceptibility is:

$$\chi_g^{3A}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} + \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{x^2 - y^2}}\left[C\left(\frac{2+y}{x}\right) - C\left(\frac{y-2}{x}\right)\right] \quad (23)$$

Region 1B. Region 1B is defined by $\mathrm{Re}[y] > x$ and $\mathrm{Re}[y] + x < 2$. This region enables the propagating graphene plasmons while inhibiting other e–h excitations. The expression for the susceptibility is:

$$\chi_g^{1B}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} + \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{y^2 - x^2}}\left[F\left(\frac{2+y}{x}\right) - F\left(\frac{2-y}{x}\right)\right] \quad (24)$$

Region 2B. Region 2B is defined by $\mathrm{Re}[y] > x$, $\mathrm{Re}[y] + x > 2$ and $\mathrm{Re}[y] - x < 2$. This region also enables the propagating graphene plasmons but includes other e–h excitations. The expression for the susceptibility is:

$$\chi_g^{2B}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} + \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{y^2 - x^2}}F\left(\frac{2+y}{x}\right)$$
$$+ \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{y^2 - x^2}}C\left(\frac{2-y}{x}\right) \quad (25)$$

Table 1 The TMD material permittivities that were used in our simulations.

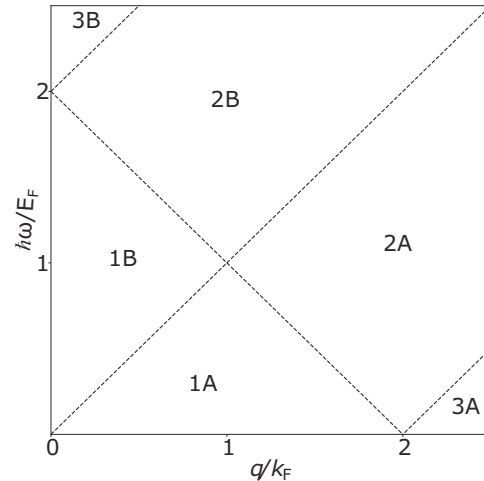| Material	In-plane static dielectric constant	Out-of-plane static dielectric constant
MoS_2	16.1	7.2
WS_2	14.5	6.6
$MoSe_2$	17.3	8.3
WSe_2	15.7	7.8

Fig. 7 Regions for the calculation of the graphene susceptibility. Regions in $(q/k_F, \omega/E_F)$-space in which the graphene susceptility is to be calculated.

Region 3B. Region 3B is defined as $\mathrm{Re}[y] - x > 2$ and it captures all of the high energy interband transitions. The expression for the susceptibility is:

$$\chi_g^{3B}(x, y) = \frac{-2}{\pi}\frac{k_F}{\hbar v_F} + \frac{k_F}{\hbar v_F}\frac{x^2}{4\pi\sqrt{y^2 - x^2}}\left[F\left(\frac{2+y}{x}\right) - F\left(\frac{2-y}{x}\right)\right]$$
$$- i\frac{k_F}{\hbar v_F}\frac{x^2}{4\sqrt{y^2 - x^2}} \quad (26)$$

Data availability

The data that support the plots and findings of this paper are available from M.K.S. (markas@dtu.dk) upon reasonable request.

References

1. Geim, A. K. & Grigorieva, I. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).
2. Novoselov, K. S., Mishchenko, A., Carvalho, A. & Castro Neto, A. H. 2D materials and van der Waals heterostructures. *Science* **353**, 6298 (2016)
3. Schmidt, P., Vialla, F. & Latini, S. Nano-imaging of intersubband transitions in van der Waals quantum wells. *Nature Nanotech* **13**, 1035–1045 (2018).
4. Basov, D. N., Fogler, M. M. & García de Abajo, F. J. Polaritons in van der Waals materials. *Science* **354**, 6309 (2016)
5. Jablan, M., Buljan, H. & Soljačić, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
6. Koppens, F. H., Chang, D. E. & Garcia de Abajo, F. J. Graphene plasmonics: a platform for strong light–matter interactions. *Nano Lett.* **11**, 3370–3377 (2011).
7. 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).
8. Kurman, Y. & Kaminer, I. Tunable bandgap renormalization by nonlocal ultra-strong coupling in nanophotonics. *Nature Physics* **16**, 868–874 (2020).

9. Andersen, M. L., Stobbe, S., Sørensen, A. S. & Lodahl, P. Strongly modified plasmon–matter interaction with mesoscopic quantum emitters. *Nat. Phys.* **7**, 215–218 (2011).

10. Takase, M. et al. Selection-rule breakdown in plasmon-induced electronic excitation of an isolated single-walled carbon nanotube. *Nat. Photonics* **7**, 550–554 (2013).

11. Latini, S., Ronca, E., De Giovannini, U., HÄbener, H. & Rubio, A. Cavity control of excitons in two-dimensional materials. *Nano Lett.* **19**, 3473–3479 (2019).

12. Törmä, P. & Barnes, W. L. Strong coupling between surface plasmon polaritons and emitters: a review. *Rep. Prog. Phys.* **78**, 013901 (2014).

13. Reithmaier, J. P. et al. Strong coupling in a single quantum dot–semiconductor microcavity system. *Nature* **432**, 197–200 (2004).

14. Scheel, S. & Buhmann, S. Y. Macroscopic QED-concepts and applications. *arXiv preprint arXiv:0902.3586* (2009).

15. Van Vlack, C., Kristensen, P. T. & Hughes, S. Spontaneous emission spectra and quantum light-matter interactions from a strongly coupled quantum dot metal-nanoparticle system. *Phys. Rev. B* **85**, 075303 (2012).

16. Raza, S., Bozhevolnyi, S. I., Wubs, M. & Mortensen, N. A. Nonlocal optical response in metallic nanostructures. *J. Phys. Condens. Matter* **27**, 183204 (2015).

17. Andersen, K., Latini, S. & Thygesen, K. S. Dielectric genome of van der Waals heterostructures. *Nano Lett.* **15**, 4616–4621 (2015).

18. Ruggenthaler, M., Tancogne-Dejean, N., Flick, J., Appel, H. & Rubio, A. From a quantum-electrodynamical lightä"matter description to novel spectroscopies. *Nat. Rev. Chem.* **2**, 1–16 (2018).

19. Flick, J., Ruggenthaler, M., Appel, H. & Rubio, A. Kohn–Sham approach to quantum electrodynamical density-functional theory: Exact time-dependent effective potentials in real space. *Proc. Natl Acad. Sci.* **112**, 15285–15290 (2015).

20. Ruggenthaler, M. et al. Quantum-electrodynamical density-functional theory: Bridging quantum optics and electronic-structure theory. *Phys. Rev. A* **90**, 012508 (2014).

21. Flick, J., Ruggenthaler, M., Appel, H. & Rubio, A. Atoms and molecules in cavities, from weak to strong coupling in quantum-electrodynamics (QED) chemistry. *Proc. Natl Acad. Scie.* **114**, 3026–3034 (2017).

22. Wang, D. S., Neuman, T., Flick, J. & Narang, P. Weak-to-strong light-matter coupling and dissipative dynamics from first principles. *arXiv preprint arXiv:2002.10461.* (2020).

23. Tokatly, I. V. Time-dependent density functional theory for many-electron systems interacting with cavity photons. *Phys. Rev. Lett.* **110**, 233001 (2013).

24. Neuman, T., Esteban, R., Casanova, D., García-Vidal, F. J. & Aizpurua, J. Coupling of molecular emitters and plasmonic cavities beyond the point-dipole approximation. *Nano Lett.* **18**, 2358–2364 (2018).

25. Jestädt, R., Ruggenthaler, M., Oliveira, M. J., Rubio, A. & Appel, H. Light-matter interactions within the Ehrenfest–Maxwell–Pauli–Kohn–Sham framework: fundamentals, implementation, and nano-optical applications. *Adv. Phys.* **68**, 225–333 (2019).

26. Kaasbjerg, K., Thygesen, K. S. & Jauho, A.-P. Acoustic phonon limited mobility in two-dimensional semiconductors: deformation potential and piezoelectric scattering in monolayer MoS_2 from first principles. *Phys. Rev. B* **87**, 235312 (2013).

27. Hinsche, N. F. et al. Spin-dependent electron-phonon coupling in the valence band of single-layer WS_2. *Phys. Rev. B* **96**, 121402 (2017).

28. Dias, E. J. C. et al. Probing nonlocal effects in metals with graphene plasmons. *Phys. Rev. B* **97**, 245405 (2018).

29. Iranzo, D. A. Probing the ultimate plasmon confinement limits with a van der Waals heterostructure. *Science* **360**, 291–295 (2018).

30. Lundeberg, M. B. Tuning quantum nonlocal effects in graphene plasmonics. *Science* **357**, 187–191 (2017).

31. Vogel, W. & Welsch, D.-G. *Quantum optics* (John Wiley, Sons, 2006).

32. Hatfield, B. *Quantum field theory of point particles and strings* (CRC Press, 2018)

33. Glauber, R. J. & Lewenstein, M. Quantum optics of dielectric media. *Phys. Rev. A* **43**, 467–491 (1991).

34. Raabe, C., Scheel, S. & Welsch, D.-G. Unified approach to QED in arbitrary linear media. *Phys. Rev. A* **75**, 053813 (2007).

35. Novotny, L. & Hecht, B. *Principles of nano-optics* (Cambridge university press, 2012).

36. Lalanne, P. et al. Quasinormal mode solvers for resonators with dispersive materials. *JOSA A* **36**, 686–704 (2019).

37. Dreizler, R. M. & Gross, E. K. *Density functional theory: an approach to the quantum many-body problem* (Springer Science, Business Media, 2012)

38. Gonçalves, P. A. D. & Peres, N. M. *An introduction to graphene plasmonics* (World Scientific, 2016)

39. Ren, J., Franke, S., Knorr, A., Richter, M. & Hughes, S. Near-field to far-field transformations of optical quasinormal modes and efficient calculation of quantized quasinormal modes for open cavities and plasmonic resonators. *Phys. Rev. B* **101**, 205402 (2020).

40. Kockum, A. F., Miranowicz, A., De Liberato, S., Savasta, S. & Nori, F. Ultrastrong coupling between light and matter. *Nat. Rev. Phys.* **1**, 19–40 (2019).

41. Schäfer, C., Ruggenthaler, M., Rokaj, V. & Rubio, A. Relevance of the quadratic diamagnetic and self-polarization terms in cavity quantum electrodynamics. *ACS Photonics* **7**, 975–990 (2020).

42. Srinivasan, K. & Painter, O. Linear and nonlinear optical spectroscopy of a strongly coupled microdiskä"quantum dot system. *Nature* **450**, 862–865 (2007).

43. Hennessy, K. et al. Quantum nature of a strongly coupled single quantum dot–cavity system. *Nature* **445**, 896–899 (2007).

44. Loo, V. et al. Optical nonlinearity for few-photon pulses on a quantum dot-pillar cavity device. *Phys. Rev. Lett.* **109**, 166806 (2012).

45. Feist, J., Fernández-Domínguez, A. I. & García-Vidal, F. J. Macroscopic QED for quantum nanophotonics: emitter-centered modes as a minimal basis for multi-emitter problems. *arXiv preprint arXiv:2008.02106* (2020).

46. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

47. Haastrup, S. et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials* **5**, 042002 (2018).

48. Cooper, V. R. Van der Waals density functional: An appropriate exchange functional. *Phys. Rev. B* **81**, 161104 (2010).

49. Enkovaara, J. E. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J Phys. Condens. Matter* **22**, 253202 (2010).

50. Chew, W. C. *Waves and fields in inhomogeneous media* (IEEE press, 1995)

51. Wasey, J. & Barnes, W. Efficiency of spontaneous emission from planar microcavities. *J. Modern Optics* **47**, 725–741 (2000).

52. Laturia, A., Van de Put, M. L. & Vandenberghe, W. G. Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: from monolayer to bulk. *npj 2D Mater. Appl.* **2**, 1–7 (2018).

53. Dias, E. J. C. et al. Probing nonlocal effects in metals with graphene plasmons. *Phys. Rev. B* **97**, 245405 (2018).

Author contributions

F.K., P.S., M.K.S., Y.K., I.K., and K.S.T. were part of the process to conceive the idea for this project. M.K.S. and Y.K. derived and implemented the MQED-DFT framework and K.S.T and I.K. supervised the project. All authors took part in drafting the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-23012-3.

Correspondence and requests for materials should be addressed to M.K.S.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints