Performance of point charge embedding schemes for excited states in molecular organic crystals

Amir Sidat; Michael Ingham; Miguel Rivera; Alston J. Misquitta; Rachel Crespo-Otero

J. Chem. Phys. 159, 244108 (2023)
https://doi.org/10.1063/5.0177278
Performance of point charge embedding schemes for excited states in molecular organic crystals

Cite as: J. Chem. Phys. 159, 244108 (2023); doi: 10.1063/5.0177278
Submitted: 20 September 2023 • Accepted: 4 December 2023 • Published Online: 27 December 2023

Amir Sidat,1 Michael Ingham,2 Miguel Rivera,2 Alston J. Misquitta,1,a) and Rachel Crespo-Otero2,b)

AFFILIATIONS
1 School of Physical and Chemical Sciences, Queen Mary University of London, London, United Kingdom
2 Department of Chemistry, University College London, London, United Kingdom

a) E-mail: a.j.misquitta@qmul.ac.uk
b) Author to whom correspondence should be addressed: r.crespo-otero@ucl.ac.uk

ABSTRACT
Modeling excited state processes in molecular crystals is relevant for several applications. A popular approach for studying excited state molecular crystals is to use cluster models embedded in point charges. In this paper, we compare the performance of several embedding models in predicting excited states and $S_1 - S_0$ optical gaps for a set of crystals from the X23 molecular crystal database. The performance of atomic charges based on ground or excited states was examined for cluster models, Ewald embedding, and self-consistent approaches. We investigated the impact of various factors, such as the level of theory, basis sets, embedding models, and the level of localization of the excitation. We consider different levels of theory, including time-dependent density functional theory and Tamm–Dancoff approximation (TDA) (DFT functionals: ωB97X-D and PBE0), CC2, complete active space self-consistent field, and CASPT2. We also explore the impact of selection of the QM region, charge leakage, and level of theory for the description of different kinds of excited states. We implemented three schemes based on distance thresholds to overcome overpolarization and charge leakage in molecular crystals. Our findings are compared against experimental data, $G_0W_0$-BSE, periodic TDA, and optimally tuned screened range-separated functionals.

INTRODUCTION
Excited states in molecular aggregates underpin several processes in biological and materials sciences. Molecular crystals, where the molecules are held by weak intermolecular interactions, are relevant for the development of optoelectronic devices, biological sensors, and pharmaceuticals, among other applications. Modeling excited states in crystalline organic materials is regularly approached by using models based either on clusters or on periodic boundary conditions (PBC). In the cluster models, the excited states are assumed to be localized on one molecule or a small aggregate, with the remainder of the crystal represented by point charges [point charge embedding (PCE)] or molecules described by a lower level of theory or a dielectric medium (continuum models). In PBC models, the system is approximated by the repetition of periodic images and the excited states can potentially delocalize over the whole crystal. Under weak light–matter interactions, light absorption from molecular crystals at room temperature results in a relatively fast localization of excitons because of the weak coupling between electronic and intermolecular vibrational degrees of freedom. For example, for a series of 13 molecular crystals of flexible molecules exhibiting solid state luminescence enhancement, the values of reorganization energies are in the range of 0.5–0.7 eV, while the exciton couplings cover the range 0.05–0.15 meV, implying fast exciton localization with transport via incoherent hopping at room temperature.1 Because of this, the utilization of embedding techniques where the excitation is localized in a central region is completely justifiable.

Modeling light-activated mechanisms and photochemical processes, such as excited state proton transfer, cis-trans isomerization, and photoinduced bond breaking/formation, requires investigation of regions of the potential energy surface beyond the Franck–Condon region, which can become very challenging in the condensed phase. The excited state potential energy surfaces are
significantly affected by the electronic structure method of choice and the employed model. While electronic structure techniques based on many body perturbation theory including methods such as Green’s function (GW) and the Bethe–Salpeter equation (BSE) can provide an accurate description of ionized and excited states, their results could be very dependent on the nature of initial orbitals used for the calculation. Additionally, their ability to explore different regions of the excited state potential energy surfaces is somehow limited and there is still much to explore in this area.

From a practical point of view, implementing point charge embedding approaches is relatively simple since most electronic structure codes allow evaluation of point charges, energies, forces and a variety of properties at different levels of theory. A good compromise to take into account the localized nature of the excitations and the long-range Coulombic interactions present in the crystal is the use of periodic embedding approaches. This is the case of Ewald embedding (EE), where a cluster model obtained from the crystal is divided into three regions (region I–III). The charges of the most external region (III) are optimized to reproduce the total electrostatic potential (ESP) in the central (I) and buffer (II) regions including the effect of the whole crystal. In practical terms, this is done by solving a series of linear equations equating the exact Coulomb and the Ewald potentials allowing to obtain the charges in region III.

Ewald embedding has been exploited for investigating the effect of defects in inorganic crystals and analyzing local properties such as nuclear magnetic resonance (NMR) parameters in molecular crystals, but only recently has it been used in the context of excited states. Based on the EE scheme, the authors of the work of Ciofini et al. proposed a self-consistent approach to consider the mutual polarization of the environment and a centrally excited molecule. Our group implemented an embedding scheme based on ONIOM (Our own N-layered Integrated molecular Orbital and Molecular mechanics) that allows for optimization of excited state minima and conical intersections using EE. All these methods are available in the Framework for Molecular Aggregate Excitations (fromage).

Time-dependent density functional theory (TDDFT) is by far the most popular method for the calculation of excited states in molecular systems, and periodic TDDFT implementations are also available. In many cases, TDDFT gives excellent results, if the correct functional is chosen. However, no matter the functional, TDDFT cannot describe some photochemical processes correctly because of its single reference nature, lack of consideration of double excitations, and wrong description of the topology of S1–S0 conical intersections. One of the advantages of adopting embedding approaches is the possibility of describing the main excited state processes using the electronic structure methods available for gas phase photochemistry, including multicorrelational methods. Even though TDDFT can be used with periodic boundary conditions, combining embedding clusters with TDDFT is a common strategy to do routine calculations with hybrids and long-range separated functionals that are still too computationally expensive in periodic approximations. In this paper, we considered periodic Tamm–Dancoff approximation (TDA), as implemented in CP2K. In contrast with TDDFT, which includes both excitations and de-excitation processes, TDA only considers excitations. It is essential to go beyond the general gradient approximation (GGA) functionals to be able to describe localized and charge-transfer excitations in organic crystals. A relevant approach presented in the work of Kronik et al. involved tuning the range-splitting parameters considering the long-range dielectric screening directly in the functional form.

The recent popularity of frozen-density approximations has expanded the toolbox of approaches available to investigate excited state processes in the condensed phase. Methods based on TDDFT-in-DFT and WF-in-DFT combining localized and periodic approaches appear to be a viable choice for studying some of the problems and applications previously mentioned. However, they are not exempt from challenges also found in PCE, such as charge leakage, overpolarization, and electron spill-out. We will highlight where these issues can become significant for PCE, as each of them becomes more problematic as the QM region approaches the point charges that it is embedded in.

This paper aims to provide some general guidelines for the use of embedding PCE techniques for the description of excited states in molecular crystals and to illustrate some of challenges commonly found in these calculations. We wish to continue our previous work on embedding methods to give a more complete assessment on the basis set dependence, testing the effects of augmented diffuse functions and the degree of ζ for the commonly used Dunning basis sets. We will also look at how diffuse and polarization functions affect the accuracy with Pople basis sets, as these are still commonly used in the field as well as in our group’s previous work. We also wish to assess the dependence on whether the point charge background is obtained from the ground or excited state and the type of electronic structure method used to describe the QM region, especially in cases where multireference effects are important. We consider these effects by comparing with selected molecular crystals from the X23 database and the calculations from Kronik’s group using optimally tuned screened range-separated functionals and G0W0-BSE as reference, as well as direct comparisons with experimental results. When overpolarization and electron spill-out become significant, we attempt to explain where this originates and give guidance on how to solve these issues practically. We implemented and tested three schemes (Zoa) to overcome overpolarization and charge leakage in molecular crystals considering distance thresholds. We focus on the methods implemented in fromage, but our conclusions are valid for other PCE methods.

**COMPUTATIONAL DETAILS**

We explore the performance of a set of embedding techniques based on point charge embedding for the prediction of S1–S0 optical gaps and other properties of interest for solid state photochemistry (Table I). We choose a set of crystals from the X23 set of molecular solids previously investigated in the work of Manna et al.

Our initial investigation of basis set dependence focused on NH3, CO2, urea, pyrazine, imidazole, benzene, and naphthalene. We consider a combination of the Pople (6-31G, 6-31G++G(d,p), 6-311G, 6-311G** and 6-311G***) and Dunning basis sets (cc-pVDZ, aug-cc-pVDZ, d-aug-cc-pVDZ, cc-pVTZ, aug-cc-pVTZ, d-aug-cc-pVTZ, cc-pVQZ, aug-cc-pVQZ, and d-aug-cc-pVQZ). We extended our analysis to triazine, uracil, and cytosine upon finding an optimal basis set (aug-cc-pVDZ and aug-cc-pVTZ) based on
TABLE I. Systems considered in this study, including the $S_1-S_0$ density difference and the experimental optical gaps (eV) used as reference. The density differences are computed at the ωB97X-D/aug-cc-pVDZ level of theory. The ground state density is shown in blue and the excited state in pink. Experimental reference data are obtained by considering the first peak at the onset of the absorption spectra as the $S_0$ energy we compare our computed results with.

<table>
<thead>
<tr>
<th>System</th>
<th>Geometry</th>
<th>Density</th>
<th>$S_1-S_0$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₃</td>
<td></td>
<td></td>
<td>6.6</td>
</tr>
<tr>
<td>CO₂</td>
<td></td>
<td></td>
<td>8.9</td>
</tr>
<tr>
<td>Urea</td>
<td></td>
<td></td>
<td>7.0</td>
</tr>
<tr>
<td>Pyrazine</td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>Imidazole</td>
<td></td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>Benzene</td>
<td></td>
<td></td>
<td>4.7</td>
</tr>
<tr>
<td>Naphthalene</td>
<td></td>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td>Triazine</td>
<td></td>
<td></td>
<td>3.7</td>
</tr>
<tr>
<td>Uracil</td>
<td></td>
<td></td>
<td>4.9</td>
</tr>
<tr>
<td>Cytosine</td>
<td></td>
<td></td>
<td>4.6</td>
</tr>
</tbody>
</table>

The geometries of the chosen systems along with example state densities and their experimental absorption energies that we use as a comparison are all shown in Table I.

The initial geometries used in this study have been obtained from the X23 set of molecular solids studied in the work of Manna et al. This database of unit cell geometries was optimized in the work of Manna et al. with periodic DFT at the Perdew–Burke–Ernzerhof (PBE)+Tkatchenko–Sheffler (PBE+TS) level of theory, where TS refers to the pairwise dispersion interaction contribution. We believe that this is valid because (1) the equilibrium $S_0$ geometry will differ only slightly from the periodic DFT optimized structure and (2) because it gives us a fair comparison with the methods used in the work of Manna et al. Using fromage, we generated clusters of molecules from the previously described unit cells such that the central QM region computed at a high level of theory is surrounded by its nearest neighbor molecules computed at the QM level of theory. The choice of which molecule from the unit cell is used in the QM region does not significantly affect the resulting excitation energies in the crystals that we have considered. This is because the molecules are similar, with similar internal coordinates, e.g., the bond lengths are within 0.001 Å of each other.

The clusters generated with a single molecule in the QM region contained a varying number of molecules as the QM region depending on the system considered (in parentheses): NH₃ (12), CO₂ (18), uracil (14), pyrazine (14), imidazole (20), benzene (16), naphthalene (12), triazine (14), uracil (20), and cytosine (16). The cluster size was chosen to include a single “solvation” shell of molecules around the central region. This allows us to consider the short-range (<4 Å) electrostatic effects. If long-range electrostatics are important, this will be accounted for in the Ewald point charge embedding calculations. In the considered crystals, the QM molecules are found in isotropic environments, i.e., the choice of molecule does not change the environment significantly. To test this, we constructed molecular clusters (QM regions) around different QM regions from the unit cell of cytosine. We then computed the excitation energy of each one by embedding in ground state charges. We found that the excitation energies (in eV) were the same to four significant figures.

The charges from the QM region are computed at the ωB97X-D level of theory with the same basis that the QM region is being computed in (PBE0 computations use PBE0 charges). Furthermore, when we consider other excited state methods, we reuse the charges obtained from (TD)-DFT derived electron densities. This has successfully been used in the groups’ previous works to model organic molecular crystals. Point charges are often derived from a ground or excited state population analysis (e.g., Mulliken and Löwdin) or derived from the molecular electrostatic potential [e.g., ESP and restrained electrostatic potential (RESP)]. We focus on RESP charges, as they do not differ significantly with respect to the basis set and capture the effects of changing density when moving to the excited state. PCE approaches have been used with all electronic structure approaches. They allow the exploration of different regions of the potential energy surface and are the best choice to explore photochemical processes in molecular crystals. The QM region in this case consists of point charges generated with different methods, outlined in Table II. For each case, we computed the charges of the...
The point charges used herein are computed for the isolated QM region molecule, then assigned to each of the QM* molecules. The geometry chosen is the same as the one that is in the QM region. The RESP charges are restrained to reproduce the quantum mechanical (QM) electrostatic potential at points around the molecule. Our previous work has shown that the use of periodic charges does not have any significant effect on the results. Since the charges are obtained from the QM region, we use a self-consistent procedure to account for the effect of the environment. The charges are computed for the QM region in the presence of the point charges of the QM* region. Those charges from the QM region are then re-assigned to the environment and the charges are recomputed for the QM region in the presence of the newly assigned point charges from the previous iteration of the self-consistent process. This is repeated until there is convergence in the summed difference of value of the charges between iterations. This sum needs to be within a convergence threshold (0.001e) before the self-consistent loop is considered converged.

The types of charges are defined as follows: EC-S_0 refers to a QM region embedded in RESP charges derived from a ground state QM calculation at the same level of theory. EC-S_1 is the same procedure except the charges are derived from an S_1 excited state QM calculation. The use of S_0 and S_1 charges allows us to explore two extreme situations: (1) The excitation is clearly localized on one molecule and the rest remains in the ground state. (2) All molecules are excited simultaneously, but only the central one is explicitly described while the other molecules are represented by point charges. Our previous studies indicate that considering ground state charges produces the most consistent results, which is in line with the significant localized excitation observed in organic molecular crystals. However, as we are testing a larger set of systems, it could be useful to evaluate if this also applies to crystals with excited states of different characters. To take into account the mutual polarization of the molecule and the environment, we also consider the excited state produced using charges obtained with self-consistent (SC) approaches. In the SC algorithms, the RESP charges are recalculated after the excited states are obtained, yielding a new set of charges, as explained in the previous paragraph. Using this SC procedure with charges from the S_0 state gives the SC-EC-S_0 approach and with S_1 charges gives the SC-EC-S_1 approach. Each of the described methods has an equivalent Ewald embedding version, where ~10000 point charges are fitted to reproduce the Ewald potential of the QM region, as discussed in Ref. 12. Reference 4 explains the original implementation in the Ewald program. These Ewald point charges are denoted by an additional “E” in the prefix of the charge type [Ewald embedded charges (EEC), self-consistent Ewald embedded charges (SC-EEC)]. All charge types were computed using Tinker in combination with Gaussian 16. A damping factor of 0.75 has been applied where it is required to give convergence of the point charges (self-consistent methods).

To assess excitonic effects caused by the interaction between molecules in the QM region, we considered the urea dimer, with 20 molecules considered in the QM* region. For the systems NH_3, CO_2, urea, and imidazole, we compute the entire spherical cluster in the QM region, consisting of 13, 19, 15, and 21 molecules, respectively. This allows us to check if the molecular density is localized to a small portion of the cluster or if the excitation is delocalized. We can then connect this with the behavior of the different types of point charges. These computations were performed at the 0B97X-D/aug-cc-pVDZ level of theory.

Both benzene and naphthalene are known to have considerable multiconfigurational character, shown by the accurate reproduction of spectra only upon computing the excited states with CASSCF. To model these molecules, we consider the π valence electrons in the CAS(6,6), CAS(12,12), CAS(10,10), and CAS(12,9) active spaces for benzene, the benzene dimer, naphthalene, and triazine, respectively. For benzene, the benzene dimer, and naphthalene, we use a state average (SA-CASSCF) of the ground state and first excited state with equal weighting. For triazine, we use a state average of the ground state and the first five excited states with equal weighting. The geometry of the benzene dimer is T-shaped and taken from the crystal phase. Multistate CASPT2 (MS-CASPT2) computations were then performed with the converged CASSCF wavefunctions in the presence of the different types of point charges previously computed at the (TD)-DFT level of theory to obtain the final excited state energies presented in this work. All MS-CASPT2 computations are done without any empirical or imaginary shift.

Solid state TD-ωB97X-D and TD-PBE0 calculations were performed using the Gaussian Plane Waves (GPWs) method of CP2K under PBCs. In the GPW methods, the LR-TDDFT (Linear Response Time Dependent Density Functional Theory) calculation is performed in a Gaussian basis set (TZVP-MOLOPT-GTH), with an auxiliary plane wave basis. The current LR-TDDFT implementation in CP2K is within the TDA; therefore, we refer to these calculations as periodic TDA going forward. Calculations were performed at the Ω-point only, and the Goedecker–Teter–Hutter (GTH) pseudopotentials were used with an energy cutoff of 500 Ry.
Calculations were accelerated using the auxiliary density matrix method (ADMM) with the cpFIT3 auxiliary basis set, necessary to simulate sufficiently large supercells with hybrid functionals. A truncated cutoff of half the smallest cell parameter was used to calculate Hartree–Fock exchange (HFX). The following supercells were used: NH$_3$ (222), CO$_2$ (222), urea (222), pyrazine (123), imidazole (121), benzene (111), naphthalene (111), triazine (112), uracil (112), and cytosine (112).

To mitigate overpolarization, we implement three simple charge redistribution schemes in fromage, based on an intermolecular distance threshold, $z_{\text{thresh}}$. In these schemes, $Z_{\text{int}}$-1, $Z_{\text{int}}$-2, and $Z_{\text{int}}$-3, point charges close to the QM/QM' boundary are zeroed and are then redistributed uniformly across the remaining point charge to maintain the overall charge. The three schemes differ in how many point charges are removed. In $Z_{\text{int}}$-1, all atoms within $z_{\text{thresh}}$, denoted as M1, are eliminated. In $Z_{\text{int}}$-2, the M1 atoms are identified as before, but additionally all atoms bonded to the M1 atoms, termed M2, are also removed. In $Z_{\text{int}}$-3, all M3 atoms and those bonded to M2 as well are removed. The $Z_{\text{int}}$ scheme differs from previous Z-scheme implementations, where the M1 atoms are joined to the QM region via a covalent bond. In $Z_{\text{int}}$, we extend such schemes to a more general intermolecular distance parameter, the connectivity of the molecule. This provides a straightforward protocol for redistributing problematic point charges in molecular crystals.

**RESULTS AND DISCUSSION**

In the following sections, we describe the results of our benchmarking of the optical gaps (S$_1$–S$_0$) with respect to experimental data, GaW$_2$-BSE, periodic TDA, and optimally tuned screened range-separated functionals. In line with the work of Manna et al., we used the experimental values obtained from the onset of absorption spectra that should provide an approximation of the optical gap of the material. The experimental data come from different experimental setups with varying conditions and errors; therefore, comparison of these values should be approached with caution. We consider the performance of charges obtained for S$_0$ and S$_1$ states. We also explore the performance of different electronic structure methods. Overall, we can divide our set of systems into two classes, strongly interacting (urea, imidazole, uracil, cytosine) and weakly interacting (NH$_3$, CO$_2$, benzene, naphthalene, pyrazine, triazine), based on their intermolecular distances in the crystal (distances <2 Å are considered to be strongly interacting).

In the considered crystals, there is an effect of packing in their excited states and their S$_1$–S$_0$ optical gaps, which originate from either electrostatic interactions or geometrical relaxation in a constrained environment. In Table S9, we show the S$_1$–S$_0$ optical gaps calculated using ωB97X-D/aug-cc-pVDZ with TDA and TDDFT for both the gas phase and periodic TDA calculations. The calculations were performed for geometries in the crystalline phase, except for the reference provided after optimization in the gas phase. The comparison between fixed geometry in the gas phase and periodic calculations reveals that for NH$_3$, urea, and imidazole, the differences are on the order of 1 eV, while for uracil and cytosine, the differences are ~0.6 eV. In systems with weaker intermolecular interactions, such as CO$_2$, triazine, and naphthalene, the deviations between periodic and gas phase calculations are around 0.2 eV. Benzene and pyrazine exhibit negligible effect in their optical gaps. However, if we consider the optimized geometry in the gas phase as the reference, even for these weak interactions, there is a shift on the order of 0.3 eV for these systems, and even larger deviations are observed for the remaining strongly interacting molecular crystals.

**EC-S$_0$: Dependence with basis set**

First, we investigated the dependence of the S$_1$–S$_0$ excitation energy with the basis set. We considered the functional ωB97X-D and several basis sets (see the section on Computational Details). The excitation energies for all the systems are shown in the supplementary material (Fig. S1). In Fig. 1, we illustrate the basis set dependence for the EC-S$_0$ method for pyrazine and imidazole.

As the basis set size is increased, there is a decrease in the deviation of the Dunning basis set excitation energies with respect to the experimental energies and the values computed in the work of Manna et al., with both pyrazine and imidazole. As augmented (diffuse) functions are added, the energy difference with respect to experimental results decreases. Furthermore, as more basis functions are used to describe each valence orbital (ξ), the difference decreases, but less so than the decrease seen when adding diffuse functions.

Pyrazine is an apolar system without hydrogen bonding and has a smaller reliance on the addition of augmented basis functions than imidazole. Convergence is reached with the aug-cc-pVDZ basis, double augmentation and going beyond the double-ξ basis makes little difference to the excitation energies. Imidazole, a polar system with hydrogen bonding (N···H), has a much larger dependence on the augmented diffuse functions, as when they are included, the excitation energy decreases by 0.4 eV (for cc-pVDZ to aug-cc-pVDZ). Convergence is achieved with the d-aug-cc-pVTZ basis.

There is a less clear basis dependence for the Pople basis sets. For pyrazine, adding polarization functions increases the difference. Unlike the Dunning basis, adding diffuse functions leads to an increase in difference with respect to the energy of the basis without diffuse or polarization functions. In imidazole, a clear dependence with basis is observed again, as the energy difference decreases in turn as polarization and diffuse functions are included (for both the 6-31G and 6-311G basis sets). In general, our calculations indicate that there is more gain in adding diffuse functions in the Dunning series than further splitting valence. The results obtained for all systems with aug-cc-pDZ are very similar to those obtained with aug-cc-pVTZ (supplementary material, Fig. S2). In terms of the Pople sets, the behavior tends to be more system dependent. However, in general 6-31++G** and 6-311++G** produce the most consistent results in comparison with the experimental references and other levels of theory. A detailed analysis of the dependence of the rest of the systems with the basis sets can be found in the supplementary material.

Figure 2 shows the S$_1$–S$_0$ gaps for the selected crystals with the largest basis set considered herein (d-aug-cc-pVQZ) with TD-ωB97X-D. Overall, the d-aug-cc-pVQZ basis produces the best results in comparison with the experimental values and other levels of theory, except in the case of imidazole, which we will discuss later, we do not find significant issues with overpolarization or charge.
spilling in TDDFT/TDA. These calculations gave us a consistent comparison point to evaluate the performance of the selected functional (ωB97X-D) for the EC-S₀ cluster model. Larger deviations from the references are found for benzene and naphthalene with EC-S₀, indicating that either the level of theory or the charge method chosen is not ideal for these compounds. The TD-SRSH method has differences comparable to EC-S₀ for urea and benzene.

As we show in the next sections, these deviations are due to the significant multireference character of these excited states and the results are improved by using multireference methods. We will explore these issues in more detail in the following sections.

Overall, the results using the charges obtained with EC-S₀ are in line with those obtained with TD-SRSH within 0.1–0.2 eV as we are only explicitly considering the excited state on one molecule. Agreement with the periodic approaches, G_0W_0-BSE and periodic TDA, is also quite good, with TDA having a tendency to slightly overestimate the gaps (maximum deviation 0.5 eV for NH₃) and G_0W_0-BSE to underestimate them (0.7 eV for CO₂). These differences can be tracked by the charges employed in the calculations that we will discuss in the next section. To reduce the computational cost without hindering the quality of the basis, we consider only the aug-cc-pVDZ and aug-cc-pVTZ basis sets in future analyses (Fig. 3).

**Dependence on choice of point charges**

In this section, we describe how the excitation energies of each system depend on the choice of point charges used to represent the embedding environment. Ideally, the results obtained with the embedding methods should provide a reasonable approximation of the periodic calculations using the same electronic structure method. To provide a more coherent comparison between the cluster and TDA periodic calculations, we perform calculations of the cluster models using the TDA method with the ωB97X-D and PBE0 functionals with the aug-cc-pVDZ basis set. Periodic calculations were performed with the same functionals, details of the basis sets can be found in the section on Computational Details. We focus herein on the analysis of the results obtained with ωB97X-D (Fig. 3), the
FIG. 3. S₁–S₀ excitation energies obtained with TDA-ωB97X-D/aug-cc-pVDZ and the different charge embedding schemes compared against those determined with periodic TDA (ωB97X-D), TD-SRSH, and G_W0-BSE (see the section on Computational Details). Results where the whole cluster is put in the QM region are also shown, but they are computed with TDDFT.

values obtained with PBE0 can be found in the supplementary material.

The TDA and TDDFT results generally have good agreement, which is often the case for absorption and emission energies, as shown in Figs. S3 and S4.⁴¹ For the ωB97X-D functional, the main differences arise in uracil and cytosine, where the TDA energies are larger than the TDDFT ones. In cytosine, the vacuum TDA energy (4.8 eV) is 0.1 eV larger than the TDDFT energy (4.7 eV), indicating that this is related to the electronic structure, potentially from the cumulative effect of de-excitations that are not considered with the TDA results. In uracil, the TDA embedded energies are ~0.2 eV above the TDDFT ones, but the vacuum energies have a negligible difference, suggesting that there is a difference that arises from embedding in TDA based charges. Again, we attribute this to the difference in method, as the charges themselves have similar values, with differences on the order of 0.001e. Overall, the best agreement between the S₁–S₀ gaps obtained with periodic TDA and cluster models is obtained with S₀ charges. We analyze the specific cases below.

In the case of NH₃, CO₂, and pyrazine, we also provide the results of the S₁–S₀ optical gaps obtained with the entire cluster including several molecules calculated with TDDFT (see the section on Computational Details). The calculations with the entire cluster should emulate the results obtained with the EC methods and give us an idea of the effect of short-range intermolecular interactions. In the case of NH₃, we also considered EC-S₀ and EEC-S₀ for the entire cluster. They are shown in Table S10 and their corresponding densities in Fig. S5.

Some of the weakly bonded systems (benzene, naphthalene, pyrazine, and triazine) are similar in that they have very little dependence on either the charge method or the basis set. This is exemplified in benzene, where the range of values for each charge method is only 0.02 eV. This is a general trend across the apolar, non-hydrogen-bonded systems. The S₁–S₀ optical gaps obtained with periodic TDA are also in very good agreement with the ones calculated for the cluster models. Computed values for all of these systems overestimate the excitation energy with respect to the experiments: benzene and naphthalene by 0.7 eV, pyrazine by around 0.4 eV, and triazine by around 1 eV. The TD-SRSH method also generally overestimates the energies. With pyrazine, it is accurate enough (within 0.2 eV of experiment), but with benzene it performs just as badly as the untuned ωB97X-D functional. We attribute this to the method used, as each of the aforementioned molecules is known to have multireference character, but the importance of explicitly accounting for this is also dependent on how dynamic correlation is accounted for.⁴²–⁴⁴ As the authors of the work of Manna et al. are able to account for these effects without using a multireference method, it is possible to describe these molecules well with single reference methods, e.g., by including the right balance of correlation and exchange in the functional chosen. To test these factors, we compare other functionals and multireference methods in a later section.

NH₃ shows interesting results with a marked dependence with the point charges employed in the calculations. The periodic results most closely match S₀ based charges, with an excitation energy (7.4 eV) in between EC-S₀ (7.1 eV) and EC-SC-S₀ (7.5 eV), suggesting that the excited state even within the periodic calculation exhibits localized behavior. However, these results deviate from the results obtained with the whole cluster that are better reproduced with the S₁ charges. The entire cluster at the QM level gives an energy of 6.2 eV, close to the EC-S₁ energy of 6.1 eV, but still underestimating the experimental energy of 6.6 eV. This initially suggests
that the excited state is delocalized throughout the cluster. Indeed, the excited state of the whole cluster is delocalized across a few molecules in the cluster, but not over the entire cluster (Fig. S5). As this is a relatively small system, we could perform computations of the entire cluster embedded in S0 charges using the EC-S0 and EEC-S0 methods. We use the S0 based charges, as we expect the excitation to be confined to the cluster (shown by the density not delocalizing completely), yielding excitation energies of 6.9 and 6.9 eV, respectively (Table S10). These values are also closer to the TDA periodic values. This indicates that while EC-S1 embedding approximates the behavior of the whole cluster in vacuum, the S1–S0 gap in the periodic system is better reproduced by using ground state charges. It is also slightly better to consider self-consistent charges in this case. These results highlight the importance of a proper balance between the size of the QM region and type of charges used for the calculations.

The S1–S0 optical gaps obtained with EC-S1 and other S1 charge schemes are more accurate for CO2 than EC-S0 in comparison with TDA periodic and the experimental values. The deviations with respect to experimental values are 0.1 and 0.3 eV, respectively. These charge methods result in a smaller difference than the G0W0-BSE methods, which underestimate the excitation energy by 0.6 eV, and are instead comparatively similar to the TD-RSH method in accuracy. There is a low dependence on whether Ewald or self-consistent charges are used, only on whether the charges are computed at the S0 or S1 level, but the difference is minimal. CO2 can be considered to be well described by both the S0 and S1 derived charges.

For the strongly interacting systems, the periodic excitation energies for urea and imidazole most closely match the S0 charges. Systems with hydrogen bonds tend to have a larger dependence on the type of charge method used. The values of their charges tend to be larger in magnitude for both the ground and excited states due to the larger differences in electronegativity between atoms. For example, urea with the aug-cc-pVDZ basis set and the EC-S0 method gives an excitation energy of 7.1 eV in good agreement with the periodic TDA value of 7.4 eV. There is some slight improvement in the EC-SC-S0 slightly improves the results predicting a value of 7.3 eV. The non-Ewald EC-S1 (6.3 eV) and EC-SC-S1 (6.3 eV) charges perform poorly in comparison with the periodic TDA calculations; however, they agree with the whole cluster calculations (6.1 eV). This behavior is similar to that observed in NH3, suggesting that S1 charges tend to provide a better estimation of the isolated whole cluster in the gas phase.

In uracil, there is little overall charge dependence. Periodic calculations (5.3 eV) are most similar to those obtained using EC-SC-S1 charges (5.3 eV), but they are still similar to the results obtained using other charge types, within 0.1 eV of each other. Both uracil and cytosine have a wide range of values that are considered to be their experimental optical gap, depending on the preparation, e.g., single crystal, anhydrous film, or as a monohydrate. The first excited state of cytosine is generally overestimated with respect to experiment (4.9 eV) by all charge types as well as the periodic results. The periodic results themselves are of similar quality to those obtained using the embedded charges. For cytosine, the S1 based charges generally perform slightly better than the S0 charges, with EEC-S1 performing the best, at an accuracy similar to that of G0W0-BSE. Overall, oB97X-D overestimates the excitation energies of the nucleotides with respect to experimental values, while PBE0 performs better but also overestimates (Fig. S4). Despite this, the results obtained by using embedding methods match well with those obtained from periodic calculations, indicating their accuracy at reproducing the solid state. It appears that the results obtained by using S0 based charges more closely agree with the periodic results, despite the expectation that the delocalized nature of the periodic calculations would correspond to excited state (S1) charge embedding.

In the next section, we discuss the case of imidazole, which shows significant dependence with the type of point charges, considering the impact of the size of the QM region. Another potential issue when using point charge embedding is the electron spill-out effect. This occurs when a diffuse electron density approaches neighboring point charges, resulting in the aforementioned overpolarization, significantly affecting the excitation energies.

**Describing imidazole and other hydrogen bonded systems**

For imidazole, there is a large dependence on the charge method used when considering just a monomer in the QM region (Fig. 4). The results with clusters embedded in S0 charges produce closer results to those obtained using periodic TDA calculations (6.6 eV), predicting an optical gap between 6.3 and 6.4 eV. We see that EC-S1 charges derived from the imidazole monomer underestimate the excitation energy by over 1 eV. Extending the system to a H-bonded dimer seems to remove this large charge dependence, resulting in excitation energies within 0.2 eV of the results presented in the work of Manna et al. and experimental values. After including a dimer in the QM region, only the EC-SC-S1 charges give a large underestimation of ~2 eV. This is due to the values of these charges themselves being unreasonable, as the self-consistent procedure yields charges with large magnitudes (>1e). We analyze these results in more detail below.

Figure 5 shows the different charges obtained by the RESP procedure for the ground and excited states for the monomer and dimer cases of the strongly interacting systems. Monomer calculations use a monomer in the QM region to obtain the charges from the RESP procedure, while dimers use a calculation of the entire dimer in the QM region during the RESP procedure. The charges reported are for the atoms involved in the hydrogen bond. We can rationalize the large underestimation observed in imidazole for the EC-S1 method based on the S1 charge values. Charges for the N-H bond in the monomer were ~0.69 and ~0.46, indicating a largely repulsive interaction. This contrasts with the dimer description, with charges of ~0.66 and 0.87, an attractive interaction. The monomer charges give a poor description of the cluster, caused by overpolarization of the density by the closely neighboring (1.74 Å) EC-S1 point charges. The effect of this overpolarization could be a cause of the erratic convergence in the self-consistent procedure to generate these charges.

The RESP procedure obtains point charges that reproduce the molecular electrostatic potential from an electron density of choice (ground or excited state). RESP charges can be sensitive to small shifts in geometry, but they can yield good results if used with an appropriate wavefunction. As the molecular electronic
density can change substantially when strong H-bonds are involved, it will be more appropriate to use the H-bonded units as a whole in determining partial changes using a method like RESP. One example of this is in the multiscale modeling of fluorescent proteins, where relevant short-range electrostatic interactions between the QM and MM (Molecular Mechanics) regions must be included to obtain accurate absorption energies. Long-range electrostatic interactions can be accurately accounted for with just point

---

**FIG. 4.** (a) Excitation energies for all the charge types for an imidazole monomer and hydrogen bonded dimer, where the environment for the monomer uses charges obtained from the monomer, while the dimer calculation uses charges from the dimer applied to the environment. This is done at the TD-ωB97X-D/aug-cc-pVDZ level of theory. Periodic TDA results were obtained with the ωB97X-D functional (see the section on Computational Details). (b) Gas phase molecule of imidazole with its excited state-ground density difference surrounded by the neighboring molecules from the crystal, showing that the gas phase $S_1-S_0$ density is diffuse enough to reach the nearest neighbor molecules. (c) $S_1-S_0$ in the crystal phase environment indicates the density is less diffuse and confined to the QM region. (d) The imidazole dimer considered in the dimer excitation energies obtained in (a).

**FIG. 5.** Displaying the orientation of each dimer considered. The point charges are derived from monomers and dimers (the entire dimer is used in the RESP procedure to obtain the charges) for urea (upper left), imidazole (upper right), uracil (lower left), and cytosine (lower right). These charges were computed at the (TD)-ωB97X-D/aug-cc-pVDZ level of theory.
charges, but as the distance between molecules decreases, more accurate descriptions are necessary to describe the charge penetration and polarization at low intermolecular distances. A less sensitive charge method could be tested but previous results with RESP charges indicate that their performance is good. Moreover, the electron density can be significantly impacted by H-bonding, which would affect the resulting charges of the RESP procedure. To account for these effects, we need to include the H-bonded molecule in the QM region, which explains why the dimer excitation energies of imidazole generally improve with respect to the monomer computations.

Figure 6 shows a comparison of the excitation energies between monomers and dimers for the other strongly interacting systems. For uracil and cytosine, moving from the monomer to the dimer for these systems has little effect on the excitation energies. We can rationalize why using a dimer in the QM region resulted in less of a change in excitation energies for the other H-bonded systems: The partial charges result in the same kinds of interactions whether obtained from the monomer or dimer, e.g., in the case of ground state imidazole, repulsive.

For uracil and cytosine, the charges retain their attractive or repulsive character even if dimers are used in the RESP procedure, which explains the smaller difference in energies (\(<\sim 0.5 \text{ eV}\)) compared to excited state imidazole (\(\sim 1–2 \text{ eV}\)) when going from the monomer to the dimer. If the charges have high magnitude and there is a change in polarity of the interaction, then the dimer energies will differ significantly from the monomer values, as we see in excited state imidazole. Exciton effects in the Kasha model can also contribute to a shift in optical gap, depending on the Coulomb interaction of dipoles of each molecule in a dimer. Exchange interactions also contribute, especially when intermolecular distances are short. As previously mentioned, for the systems we are considering, these effects are considered to be small but must generally be accounted for if accurate excitation energies are desired. We can only account for this interaction in the QM region, by inclusion of dimers (larger number of molecules). It is difficult to know a priori what the correct choice of charge is. However, for most systems, the use of ground state charges and the EC-S0 models provides good agreement with the periodic calculations and considering either self-consistent or Ewald can slightly improve the results depending on the system. The dimer and whole cluster computations can confirm the nature of the excited state for the interacting system, allowing us to justify the choice of point charge used. Moreover, the packing of molecules itself is

![Image of Figure 6](image_url)
related to the intermolecular interactions present. Some systems will suffer more from issues with descriptions of the environment, especially polar systems that contain strong intermolecular interactions.

Performance of different functionals and post-Hartree methods

In this section, we compare different levels of theory. We compare the use of the PBE0 functional, CC2, and multireference methods (CASSCF and CASPT2) to capture static and dynamic correlation. Figures 7 and 8 show a comparison of these methods for the six systems that did not have accurate excitation energies with ωB97X-D. We have only included the non-self-consistent charge variants in these figures, as they are not available for the CC2 and CASSCF based methods.

The PBE0 functional is a global hybrid that has no reliance on empirical parameters, while ωB97X-D is a range-separated hybrid that adds a varying amount of Hartree-Fock exchange depending on the value of the range-separation parameter ω. This is normally adjusted to reproduce the negative orbital energies that correspond to the ionization potential and electron affinities in the gas phase. However, there is no guarantee that adjusting these properties for the gas phase will result in improved results for the solid phase, which may explain why PBE0 seems to perform as accurately as the optimally tuned TD-SRSH method shown in gray (Fig. 7). Furthermore, the systems we have studied do not contain the low energy charge-transfer states sometimes encountered with global hybrids. This means that the range-separated treatment of the exchange will not be as relevant in the description of excited states for these systems (reflected in PBE0’s similar behavior to ωB97X-D). CC2 is normally accurate for the excited states of single reference systems, but its accuracy could be disrupted by potential electron spill-out effects and how much more diffuse the CC2 density is. Both functionals overestimate the excitation energies of uracil, with slightly better performance in the case of S1 charge methods.

CC2 gives mixed results for the excitation energies. With uracil, EC-S0 shows good agreement with G0W0-BSE and experimental values, also giving results similar to those of PBE0, while giving values that are underestimated in relation to the results obtained...
by using EC-S$_1$. CC2 performs poorly for triazine and cytosine with both charge types plotted, exhibiting severe overestimation and underestimation, respectively. Ewald charges with CC2 lead to inaccurate results, so they have not been plotted. This is due to the more diffuse nature of the CC2 electron density, resulting in overpolarization as has been previously described (see also Sec. S5). In the case of EEC-S$_1$, the excitation energies are negative, which is normally indicative of a single reference determinant not being an adequate description of the excited state. This should not be the case as uracil has previously been described with single reference methods effectively. If the excited state energy is stabilized below the ground state and it is not due to the single reference determinant being the issue, then we expect that the interaction between the charges and the excited state electron density is causing it. This indicates electron spill-out from the QM region and results in the excited state energy being unreasonable.

Cytosine shows a similar pattern to that of uracil for the DFT functionals, with PBE0 and the EC-S$_1$ charges performing slightly better than EC-S$_0$. CC2 significantly underestimates the excitation energies for both charge embedding methods, which we again attribute to overpolarization. CC2 might be more susceptible to overpolarization than TD-DFT as CC2 densities are more diffuse, as can be seen in Fig. 9. One may think that these issues could be solved by including a dimer in the QM region, similar to imidazole. However, the origin of overpolarization here is from diffuse parts of the density interacting with the surrounding point charges. This issue is not solved by moving to a dimer, and we still observe overpolarization, as mentioned in the last section. We attempt to mitigate overpolarization by redistributing point charges close to the QM/QM$^*$ boundary in the next section.

For pyrazine, PBE0 performs more accurately than ωB97X-D, giving closer absorption energies to experimental values by around 0.2 eV. In the case of CC2, there is further improvement in the energies, giving values very close to the experimental reference of 3.8 eV. Gauss-W0-BSE values are underestimated here more than either of the above methods, giving an absorption energy of 3.5 eV. The embedded PBE0 calculations give similar results to those of periodic TDA as previously mentioned (3.9 eV).

The main impact on excitation energies for the aromatic systems (benzene, naphthalene, triazine) arises from the use of multireference methods. For any system where there are double or higher excitations, it is easier to incorporate the static correlation with a multireference method, e.g., SA-CASSCF and multistate CASPT2.11,12

Both benzene and naphthalene are systems that contain significant multireference character. The excitation energies for these systems show little dependence on the charge type used for the embedding, as might be expected for apolar systems. Figure 8 shows that we obtain excitation energies within 0.2 eV of the experimental values with the multireference CASPT2 wavefunction, compared to the 0.6–0.7 eV overestimation obtained with the single reference ωB97X-D approach.

MS-CASPT2 recovers the expected excitation energies very well for each charge type. Increasing the QM region for benzene to a dimer further reduces the excitation energies closer to the Gauss-W0-BSE results and experiment, suggesting that even in this non-H-bonded system, accounting for additional interactions between the molecules does result in more accurate excitation energies. This could be a result of accounting for the increased dispersion energy, attributing to the stability of benzene dimer complexes, for example, see the work of Sherill and Tsuzuki et al.50,51

For triazine, we see little dependence on the choice of embedding charge method. ωB97X-D overestimates the excitation energy by around 1 eV, and so does PBE0, albeit to a slightly smaller amount of ~0.8 eV. Interestingly, CC2 also leads to an overestimation of 1.8 eV with the S$_0$ charges and ~1.5 eV with the S$_1$ charges. This is possibly due to the effect of overpolarization discussed above. The MS-CASPT2(12,9) computations accurately describe the excited state, especially with the excited state charge methods, yielding differences of <0.1 eV from the experimental references. The performance of CASPT2 is also similar to that of Gauss-W0-BSE. Note that despite the structural similarity of triazine and pyrazine there is a stark difference in the multireference contribution to the excitation energies of these molecules.

We can gauge the effect of overpolarization on the CC2 energies by calculating the excited state energies with a less diffuse basis.
but with the same point charge background. If the density itself is too diffuse, no matter what the point charge values are, it is likely that the density will get too close to the point charges in the QM' region, leading to overpolarization (shown in Fig. 9). We recomputed the excitation energies of cytosine obtained with CC2, without diffuse functions (with the cc-pVDZ basis instead of aug-cc-pVDZ). To check that the quality of the basis is not a significant issue, we computed the vacuum energies obtained with each basis, which gave energies of 4.6 and 4.4 eV for cc-pVDZ and aug-cc-pVDZ, respectively. This indicated that the change in basis results in a 0.2 eV difference from increasing the size of the basis and gives us a baseline for what the effect of increasing the virtual space is. We then recomputed the EC-S₁ charge embedding results with the cc-pVDZ basis, giving an absorption energy of 5.2 eV, similar to the results obtained with ωB97X-D, while aug-cc-pVDZ gave an absorption energy of 3.4 eV, largely underestimating the experimental values. Since the quality of the smaller basis is not significantly detrimental to the estimation of experiment values compared to the difference caused by the point charges, it indicates that the underestimation from CC2 is down to overpolarization. It is more important for CC2 than DFT based methods to include some kind of exchange contribution to keep the density confined or to allow for the geometry of the surroundings to adjust to the density accordingly. CASPT2 is necessary to accurately describe triazine, benzene, and naphthalene, but CC2 is enough with certain charge types to describe pyrazine, uracil, and cytosine.

Reducing overpolarization with the Z_{int} scheme

Our CC2 calculations on cytosine illustrate how PCE calculations are susceptible to overpolarization by point charges close to the QM/QM' boundary. This is problematic when the charge density is diffuse due to either the method or basis set, or both. A poor choice of model region may excessively truncate the electronic structure of the crystal, resulting in over polarisation by the environment that can be improved by selecting a larger QM region. For instance, in our EC-S₁, TDDFT calculations on imidazole, using a dimer model corrects the ostensible overpolarization artifacts. In the case of cytosine, however, significant overpolarization is present in CC2/aug-cc-pVDZ calculations of both the monomer (3.4 eV) and the dimer (3.0 eV). To address the issue of overpolarization, we implement the new $Z_{int}$-1, $Z_{int}$-2, and $Z_{int}$-3 schemes in fromage, to redistribute problematic point charges close to the QM/QM' boundary evenly across the remaining point charges (see Computational Details). We investigate this both for CC2 and TDA, where overpolarization was problematic in only the former. The distance threshold controls the number of M1 charges removed; therefore, we increase this from 1.8 Å, where just two point charges are removed, up to 3.0 Å, where 21 point charges are identified.
In Fig. 10, we observe that the Z_{int}-1 and Z_{int}-2 schemes can have both a positive and negative impact on the excited state energies. In the CC2 calculations, modification with the Z_{int}-2 scheme reduces overpolarization and provides S_1 energies (4.5–4.7 eV), at a threshold of 2.2–2.7 Å, that are in line with the experimental value (4.6 eV). This is a significant improvement relative to the overpolarized EC-S_0 calculation (3.4 eV) and also to the vacuum energy (4.4 eV). Of course, at a low threshold (1.8–1.9 Å), the results are in line with those obtained with the overpolarized EC-S_0 calculation, and beyond 2.7 Å, the quality becomes poor, with the convergence problems arising at a threshold of 3.0 Å. As such, these results are an encouraging indication that more sophisticated redistribution schemes may be a fruitful approach in reducing overpolarization from PCE. In contrast, in the case of Z_{int}-1 CC2 calculations, the results were generally poor, resulting in significantly underestimated energies between 1.8 and 2.9 Å. At a threshold of 3.0 Å, when a significant number of point charges close to the model region have been removed, the S_1 energies unsurprisingly come close to that of the vacuum calculation. Calculations using the Z_{int}-3 scheme resulted in convergence issues. Careful testing is therefore recommended when using this method.

We also performed TDA calculations using the Z_{int} scheme, as a reference system in which overpolarization is not present. Here, we use the S_1 energy obtained using periodic TDA (5.4 eV) as the solid state reference due to the inherent uncertainty in experimentally obtained values. Overall, in the TDA calculations, we see the Z_{int} scheme negatively affects the excited state energies. For the Z_{int}-1 scheme, all energies are below both experimental and vacuum values (4.7 eV), although at a high threshold it appears to tend toward the vacuum calculation. A similar situation is observed in the case of Z_{int}-2, with a notable exception occurring at a threshold of 1.8–1.9 Å in which the energy obtained (5.1 eV) is improved relative to the vacuum level, but still worse than the results obtained using the unmodified EC-S_0 calculation. However, in this case, the EC-S_0 calculation result is in good agreement with the periodic reference values, indicating that the PCE cluster model is already an excellent approximation to the solid state. This is often the case for molecular crystals, where the QM/QM’ boundary does not traverse any covalent bond. Consequently, redistributing point charges in molecular crystals proves to be a negative perturbation in the absence of overpolarization. In a future work, we focus on the implementation of improved schemes for crystals that redistribute charge in a less extreme manner, specifically those that preserve the local dipole moments.

**Effect of geometry**

One advantage of using embedding methods based on PCE, in contrast to methods that do not explicitly consider the structure of the environment, such as continuum models, is the ability to capture how specific interactions with the environment can affect the excited state relaxation. This section describes the results obtained with the full ONIOM QM:QM’ scheme, to describe photoemission for systems that depend highly on the charge background (imidazole) and systems that do not depend on it (pyrazine and benzene) to show the capabilities of point charge embedding.

Geometry optimizations were performed to obtain the equilibrium geometries for the S_0, S_1, and T_1 states. The purpose of this is to check our assumptions of the absorption energy, that the optimization procedure makes little difference to the resulting excited state energies, thus validating our previous results computed without any optimization. It also allows us to test how our embedding approaches behave in the optimization of other (excited) states of interest. Geometry optimization allows us to avoid these difficulties and to obtain accurate geometries that reflect experimental excited state properties, allowing us to test our embedding methods. Ideally, we would like to compare results for geometries determined via experiment, to eliminate any dependence on the geometry optimization procedure. However, excited state reference geometries are more difficult (requiring advanced spectroscopy) to obtain than the ground state geometry, which can more simply be obtained via x-ray diffraction.
The results for imidazole, pyrazine, and benzene are shown in Table III. We also optimized S₁ in the presence of S₀ derived charges and T₁ in T₀ charges, testing the effect of charge background on the optimization, representing the case where the excitation is delocalized on the molecular cluster. These are labeled as “Excited state charges” in Table III. For benzene, we also obtain the S₀/S₁ minimum energy conical intersection instead of T₁.

Pyrazine’s phosphorescence is accurately described within 0.1 eV for both charge types. The lack of charge dependence here corroborates with the previous values obtained for absorption. The absorption energy for the fixed experimental crystal geometry is within 0.02 eV of the post-optimization value, indicating that the original geometry is reasonable. While a fluorescence energy of pure pyrazine in the crystal is to our knowledge unavailable, there are spectroscopic data for pyrazine in a benzene matrix. This preparation gives absorption, fluorescence, and phosphorescence energies of 3.8, 3.6, and 3.2 eV, respectively. The phosphorescence energy (3.2 eV) closely matches that of Ref. 56 (3.3 eV), we can assume that these energies are reasonable, and as such our computed fluorescence energy of 3.9 eV is also in close agreement with the experimental fluorescence energy of pyrazine in a benzene matrix. For imidazole, the absorbance energy is overestimated by 0.4 eV, similar to the frozen geometry case, with only a 0.04 eV difference before and after optimization. It is clear that optimization is not enough to obtain adequate excitation energies, and use of a dimer in the QM region is mandatory as seen earlier in Fig. 4.

To test the effect of including the H-bonded dimer in the QM region, we optimized the imidazole monomer and the dimer with the EC-S₀ and EC-S₁ charge methods. We find that if we include a dimer in the QM region, we obtain energies of 6.2 for both EC-S₀ and EC-S₁, bringing the absorption within 0.2 eV of experimental values, giving values similar to the non-optimized results presented previously. This shows how effective including the dimer is in improving the excited state energies, and also shows that optimization of the S₀ state is less important than that of excited states.

We now use the ONIOM(QM:QM') model to optimize the S₀, S₁, and S₀/S₁ conical intersection geometries for benzene in the crystal phase, with the aug-cc-pVDZ basis. Multiconfigurational methods are necessary to describe benzene due to the large amount of static correlation needed to describe its excited states, a result of its degenerate orbital structure. These results with benzene serve as an example to show how useful the methods based on point charge embedding could be to describe excited state structures and conical intersections.

Figure 11(a) shows the geometries obtained in the solid phase and their energies. The obtained S₀ geometry gives an absorption energy of 4.9 eV, very close to the one obtained without optimization. For the conical intersection, the half-boat structure obtained in the crystal phase matches the one determined in the work of Blancafort et al. in the gas phase. In their investigation of the S₀/S₁ conical intersection, they found this geometry to be the lowest energy of a range of geometries along the crossing seam. This indicates that the space available in the crystal [Fig. 11(b)] is enough to allow for the same geometry to form in both phases. This corresponds to the fact that benzene has larger intermolecular bond distances with weaker electrostatic interactions, allowing for large shifts in geometry without restriction. The predicted fluorescence energy is difficult to connect to experiment due quenching in the crystal phase. Spectra obtained for benzene in water give absorption and fluorescence values of 4.8 and 4.4 eV, respectively. The former is close to the computed crystal absorption (4.9 eV) and the latter corresponds to the computed fluorescence (4.5 eV). In
molecular crystals, in addition to excimer quenching, quenching can also occur via conical intersection.\(^{1,2}\) This is shown by the close \(S_1\) and \(S_1/S_0\) intersection energies, facilitating competition between fluorescence and non-radiative decay. As we obtain similar energies in the solid phase, we would predict that this deactivation channel could also be in the crystal. The geometry optimization procedure in the solid state is robust enough to allow for the formation of geometries with significant changes to bond lengths and angles, as shown with the half-boat structure.

### CONCLUSIONS

In this paper, we have explored the performance of different PCE embedding techniques for the investigation of \(S_1\)–\(S_0\) gaps for a range of systems with a variety of excited state methods. We focus on the use of QM regions of one molecule and also analyze the effect of increasing the size of the QM region to the dimer and larger monomers. As reference, we have considered excited states obtained from G0W0–BSE, TD-SRSH, where our calculations have used periodic TDA and experimental values. Our calculations indicate that PCE calculations can provide accurate values for \(S_1\)–\(S_0\) energy gaps in molecular crystals that can also approximate results from periodic, large QM region calculations and experiments.

From the analysis of basis set tests, the Dunning basis sets have a clear convergence as augmented diffuse functions are added. Our calculations indicate that, for most systems, the addition of diffuse functions yields a greater benefit than further splitting of valence. Although the predicted \(S_1\) excited states showed limited improvement when transitioning from aug-cc-pVDZ to aug-cc-pVTZ, the inclusion of diffuse functions was crucial for aligning the predictions with experimental results and other levels of theory. In the case of the Pople basis sets, 6-31++G** and 6-311++G** produce similar results.

Dimer and whole cluster calculations can help confirm the nature of the excited state for the interacting system. Overall, using \(S_0\) charges tends to provide results closer to the results obtained using periodic calculations and further improvements can be obtained by using Ewald schemes. Ewald charges on their own generally show either a similar performance or an improvement to the excited state energies when compared to the standard charges, so should be used when possible, as they do not result in significantly increased computational cost. The imidazole monomer is an exception, as Ewald charges without self-consistency underestimate the excitation energies (>1 eV). This problem is solved when going to the imidazole dimer and as such is not attributed to Ewald charges but to overpolarization at short range. For the tested systems, the calculations using \(S_1\) agree better with the simulations of the whole clusters. However, if these whole clusters are further embedded in point charges, the results are now in better agreement with the results obtained using periodic calculations.

Using a dimer instead of a monomer in the QM region is recommended to better capture the effect of intermolecular interactions and exciton couplings. If the molecule of interest has hydrogen bonding, using a dimer in the QM region can help in improving the description of excited states and stabilizing the dependence on the charge type (see Fig. 6). This comes with increased computational cost, so should only be used if there is hydrogen-bonding and/or other short-range interactions.

Depending on the system, we will need to account for other intermolecular interactions that are present, as the point charge embedding will only help obtain the electrostatic interaction. If a system has significant induction or dispersion contributions, their effect is included with the use of QM region computation. \(\omega\)B97X-D will include the exchange–correlation as well as an empirical dispersion contribution. This cannot be obtained with just point charges, especially for apolar systems. We require polarizable embedding methods or other methods to describe these additional intermolecular interactions.\(^{3,9}\) In the systems examined, we do not observe any significant non-electrostatic effect on the excitation energies.

\(\omega\)B97X-D generally overestimates excitations for conjugated systems, a known effect of the functional.\(^{10}\) Moreover, overestimations could be related to how we have not optimized the range-separation parameter \(\omega\). Reference 61 shows that this can

---

**FIG. 11.** Potential energy plot of benzene, geometries were optimized with CASSCF/aug-cc-pVDZ (See Computational Details). The CASPT2 energies at the \(S_0/S_1\) conical intersection are an average of the \(S_0\) and \(S_1\) energies. The original CASPT2 energy gap for the two states at this geometry is 0.2 eV.
lead to significantly more accurate excitation energies in conjugated systems (pentacene). However, this explanation does not explain why the optimally tuned TD-SRSH method does not correctly describe benzene, which requires the use of multiconfigurational methods. It is likely that in addition to optimal tuning, either the static or dynamic correlation is incorrectly described by the DFT functional, leading to large overestimations in the absorption energy. PBE0 was found to give better performance than \( \omega B97X-D \), but a range-separated hybrid like \( \omega B97X-D \) is necessary to give the correct description of charge-transfer states. In this case, low energy charge-transfer states were not found. It is important to check for charge-transfer states before using global hybrids without range separation.

Our calculations indicate that the selection of the electronic structure method is generally more important to achieve good accuracy than the selection of charges to represent the environment. As mentioned before, using ground state charges seems to provide a good compromise. Use of CC2 or CASPT2 depends on the importance of larger than double excitations in the excited state wavefunction of interest, where TDDFT performs worse. The cheaper method will generally be CC2, but CC2 suffers from electron spill-out and related issues due to the diffuse electron density generated (as seen in cytosine). If overpolarization and higher excitations are not pertinent to the system, then use CC2. Otherwise, CASPT2 is mandatory for any multireference system, as shown with benzene, naphthalene, and triazine. Use of a dimer here can help improve the quantitative accuracy (see Fig. 8), through a more accurate description of the short-range interaction energy components and dynamic correlation.

We have found that CC2 can be prone to overpolarization, especially in triazine, uracil, and cytosine for Ewald based charges. Using CC2 here will almost always require some form of exchange interaction to confine the generally more diffuse electron density produced by the method, to prevent the effects of electron spill-out and overpolarization. We implemented the new \( Z_{\text{int}}-1 \), \( Z_{\text{int}}-2 \), and \( Z_{\text{int}}-3 \) schemes from fromage; these schemes based on intermolecular thresholds seem to help solve the problem for the problematic CC2 calculations. In the case of cytosine, we found that the \( Z_{\text{int}}-2 \) scheme produced reasonable \( S_1-S_0 \) energies while the other two approaches did not. Our comparison with TDA periodic calculations did not indicate significant overpolarization issues for either TDA or TDDFT for the cluster models used to describe the molecular crystals. We are working to further develop and test these methods and the use of different approaches to mitigate overpolarization in QM:QM’ calculations.

We can apply the various embedding approaches in conjunction with the optimization of critical points along the potential energy surface, to give access to computed emission values. We use this to compute emission energies for pyrazine, imidazole, and benzene. We obtain particularly accurate results with pyrazine. With benzene, we obtained conical intersection geometries in the solid state, which are similar to the ones in the gas phase, indicating a similar available mechanism of non-radiative decay. While our primary focus has been on calculating the energy gaps for the first excited state (\( S_1 \)), our analysis of the excited state in various molecular crystals leads us to believe that PCE models can also adequately describe higher excited states, given the use of an appropriate set of charges.

**SUPPLEMENTARY MATERIAL**

Additional data is included in the supplementary material.

**ACKNOWLEDGMENTS**

This research has been supported by the Leverhulme Trust (Grant No. RPG-2019-122). We utilized Queen Mary’s Apocrita HPC facility, supported by QMUL Research-IT and the ARCHER UK National Supercomputing Service (Grant No. EP/X035859/1) via the Materials Chemistry Consortium and the Molecular Modeling Hub for computational resources, MMM Hub, which is partially funded by EPSRC (Grant No. EP/T022213/1). A.J.M. and R.C.-O. acknowledge receiving funding for the PHYMOL program (HORIZON-MSCA-2021-DN-01, Project No. 10103474).

**AUTHOR DECLARATIONS**

**Conflict of Interest**

The authors have no conflicts to disclose.

**Author Contributions**

**Amir Sidat**: Data curation (lead); Formal analysis (lead); Methodology (lead); Software (equal); Validation (equal); Visualization (lead); Writing – original draft (lead). **Michael Ingham**: Data curation (supporting); Formal analysis (equal); Investigation (equal); Methodology (supporting); Software (equal); Writing – review & editing (equal). **Miguel Rivera**: Conceptualization (lead); Formal analysis (equal); Funding acquisition (lead); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Writing – review & editing (equal). **Alston J. Misquitta**: Conceptualization (lead); Formal analysis (equal); Funding acquisition (lead); Methodology (equal); Project administration (equal); Resources (lead); Supervision (equal); Writing – review & editing (equal). **Rachel Crespo-Otero**: Conceptualization (lead); Data curation (supporting); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (supporting); Project administration (lead); Resources (lead); Supervision (lead); Writing – review & editing (lead).

**DATA AVAILABILITY**

The data that support the findings of this study are available within the article and its supplementary material.

**REFERENCES**


