Many-body GW calculations with very large scale polarizable environments made affordable: A fully ab initio QM/QM approach

David Amblard, Xavier Blase, Ivan Duchemin

Many-body GW calculations with very large scale polarizable environments made affordable: a fully *ab initio* QM/QM approach

David Amblard,[1] Xavier Blase,[1] and Ivan Duchemin[2]

[1)] *Univ. Grenoble Alpes, CNRS, Inst NEEL, F-38042 Grenoble, France*

[2)] *Univ. Grenoble Alpes, CEA, IRIG-MEM-L_Sim, 38054 Grenoble, France*[a)]

(Dated: 24 October 2023)

We present a many-body GW formalism for quantum subsystems embedded in discrete polarizable environments containing up to several hundred thousand atoms described at a fully *ab initio* random phase approximation level. Our approach is based on a fragment approximation in the construction of the Green's function and independent-electron susceptibilities. Further, the environing fragments susceptibility matrices are reduced to a minimal but accurate representation preserving low order polarizability tensors through a constrained minimization scheme. This approach dramatically reduces the cost associated with inverting the Dyson equation for the screened Coulomb potential W, while preserving the description of short to long-range screening effects. The efficiency and accuracy of the present scheme is exemplified in the paradigmatic cases of fullerene bulk, surface, subsurface, and slabs with varying number of layers.

[a)] Electronic mail: ivan.duchemin@cea.fr

I. INTRODUCTION

The description of the electronic properties of a quantum subsystem embedded in a polarizable, or dielectric, environment (a molecular interface, a solvant, an electrode, etc.) remains a central issue in many fields pertaining to solid-state physics, chemistry or biology. Starting from the historical image charge models for electronic distributions close to a metallic surface or within a dielectric cavity,[1,2] the need to describe the response of a polarizable environment to a charged (photoemission) or neutral (optical) excitation in a specific subsystem is still triggering significant developments to combine accuracy with numerical efficiency. In particular, the stabilization of an added hole or electron by the induced electronic rearrangements in a surrounding polarizable environment can be as large as several electronvolts. This so-called polarization energy, together with the additional effects of the electrostatic environment and wavefunction delocalization, strongly renormalize the electronic properties. In many situations, the environment is a complex, potentially infinite, system that cannot be fully described at the same quantum level as the subsystem of interest.

Conceptually close to the historical models of image charges, the polarizable continuum model (PCM)[3,4] considers a quantum subsystem located in a cavity carved into a medium described by an homogeneous macroscopic dielectric constant. As a more expensive alternative, discrete polarizable models, where atoms are described as polarizable centers, allow for a more realistic description of screening inhomogeneities at short range in response to an electronic excitation in the quantum subsystem.[5–9] Polarization energies converging slowly with environment size, thousands of polarizable centers may be needed in order to enter a regime where long-range extrapolation can be achieved on the basis of the calculated values. This comes as a challenge to fully *ab initio* approaches, triggering in practice the description of the environment at a semiclassical empirical level. In such semi-empirical discrete approaches, labeled generically QM/MM, or QM/MMpol to emphasize the polarizable nature of the environment, atoms are provided with effective polarizabilities that reproduce the correct molecular polarizability tensor and/or the macroscopic dielectric tensor of the material.[10,11]

Concerning the quantum mechanical formalism used to describe electronic excitations in the central subsystem, a specific class of many-body perturbation theories, the *GW*[12–16] and Bethe-Salpeter equation (BSE)[17–21] formalisms for the study of charged and neutral electronic excitations, have been recently combined with polarizable models of environment, both at the continuum[22–24]

and discrete[25–28] levels. Indeed, while recent studies demonstrated that *GW* calculations with cubic or even lower scaling could be achieved,[29–40] the slow convergence of electrostatic and dielectric (screening) effects with respect to system size forbids a brute force treatment of complex environments within such approaches.

The cost associated with the building of the irreducible electronic susceptibility is usually the bottleneck in time-dependent density-functional theory (TD-DFT) and *GW* calculations. As a cure, fragments approximations can dramatically reduce such a cost by neglecting wavefunction overlaps between weakly interacting subsystems. These fragment or subsystem approaches have been recently implemented at the *GW* and BSE levels in the case of systems presenting weakly interacting subunits, including molecular systems,[41–46] interfaces,[47–50] 2D materials,[51,52] but also in the less obvious case of covalent 2D systems.[53] It remains that obtaining the screened Coulomb potential W from the irreducible susceptibility, requiring a matrix inversion, restricts the number of fragments that can be dealt with at the fully *ab initio* level. As such, the largest fragment-based *GW* calculations were obtained for a system containing about 800 benzene molecules (4800 non-H atoms) in the case of molecular systems, an already remarkable achievement.[43] Similarly, we recently used a fragment *GW* approach to partition a multilayer *h*-BN system in up to 259 *h*-BN fragments containing 66 non-hydrogen atoms each.[53] The limiting factor to the simulation of larger systems was then the inversion of the Dyson equation to obtain W.

In the present study, we introduce and assess a fully *ab initio* scheme for embedded *GW* calculations with hundreds of thousand atoms in the environment. Besides adopting the fragment approximation, we search for an efficient low-rank representation of the susceptibility matrix associated with each fragment, projecting them on-the-fly onto a minimal polarization basis preserving the dipolar, quadrupolar, etc. fragments polarizability tensors. As a result, the size of the Dyson equation for W is dramatically reduced, while preserving the accuracy for the polarization energies at the meV level. We explore the trade-off between accuracy and efficiency in the case of a fullerene crystal, both in the bulk, surface, subsurface and few-layers-slab limits.

II. THEORY

We very briefly outline the *GW* formalism, directing the reader to thorough reviews and books for a more detailed account on the subject.[54–59] We further introduce our embedding scheme associated with the definition of the environmental screening, or reaction field. Finally, we describe our fitting

scheme allowing to dramatically reduce the size of the fragments dielectric matrix expressed in an effective polarization basis that preserves short to long-range screening effects.

A. The *GW* formalism

Departing from the use of the electronic density in DFT, the *GW* formalism takes as a central variable the time-ordered one-body Green's function built from input $\{\varepsilon_n, \phi_n\}$ Kohn-Sham eigenstates, namely:

$$G(\mathbf{r}, \mathbf{r}'; \omega) = \sum_n \frac{\phi_n(\mathbf{r})\, \phi_n^*(\mathbf{r}')}{\omega - \varepsilon_n + i\eta \times \mathrm{sgn}(\varepsilon_n - E_F)} \tag{1}$$

where η is a positive infinitesimal and E_F the Fermi energy. Relying on perturbation theory to low order in the screened Coulomb potential W, the energy-dependent exchange-correlation self-energy $\Sigma(\mathbf{r}, \mathbf{r}'; E)$ can be approximated by the *GW* operator under the form:

$$\Sigma^{GW}(\mathbf{r}, \mathbf{r}'; E) = \frac{i}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}\omega\, e^{i\eta\omega}\, G(\mathbf{r}, \mathbf{r}'; E + \omega)\, W(\mathbf{r}, \mathbf{r}'; \omega) \tag{2}$$

with v the bare Coulomb potential and W the dynamically screened Coulomb potential built within the random phase approximation (RPA):

$$\begin{aligned} W(\mathbf{r}, \mathbf{r}'; \omega) = {} & v(\mathbf{r}, \mathbf{r}') \\ & + \int \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{r}_2\, v(\mathbf{r}, \mathbf{r}_1)\, \chi_0(\mathbf{r}_1, \mathbf{r}_2; \omega)\, W(\mathbf{r}_2, \mathbf{r}'; \omega). \end{aligned} \tag{3}$$

Such an equation adopts a self-consistent Dyson-like form that needs to be inverted once the independent-electron susceptibility has been built from Kohn-Sham one-body eigenstates:

$$\begin{aligned} \chi_0(\mathbf{r}, \mathbf{r}'; \omega) = {} & \\ & \sum_{m,n} (f_m - f_n) \frac{\phi_m^*(\mathbf{r})\, \phi_n(\mathbf{r})\, \phi_m(\mathbf{r}')\, \phi_n^*(\mathbf{r}')}{\omega - (\varepsilon_n - \varepsilon_m) + i\eta \times \mathrm{sgn}(\varepsilon_n - \varepsilon_m)} \end{aligned} \tag{4}$$

with $\{f_{m/n}\}$ level occupation numbers. The cost of calculating this independent-electron susceptibility grows as $O(N^4)$ with respect to the number of electrons N in the system. Equivalently, dropping the space variables for compactness, the Dyson equation can be formulated as:

$$W(\omega) = v + v\,\chi(\omega)\,v, \tag{5}$$

with χ the RPA interacting susceptibility:

$$\chi(\omega) = \chi_0(\omega) + \chi_0(\omega)\, v\, \chi(\omega). \tag{6}$$

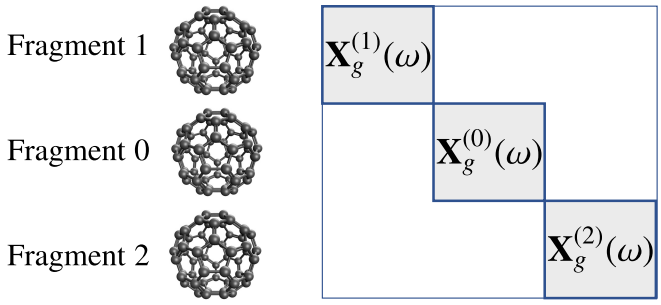


FIG. 1. Symbolic representation of the fragment approximation. The standard fragment approximation on the independent-electron susceptibility (the $\mathbf{X}_0^{(\mathrm{I})}$ blocks in Eq. 9) can be strictly reformulated by constructing first the block diagonal matrix made of the gas phase interacting susceptibility of the isolated fragments (the $\mathbf{X}_g^{(\mathrm{I})}$ in Eq. 10). It is such an object that we represent here. The resulting fully interacting susceptibility is non-diagonal due to the coupling of the blocks by the off-diagonal $\mathbf{V}^{(\mathrm{I}\neq\mathrm{J})}$ Coulomb matrix elements (see Eq. 10).

The knowledge of Σ^{GW} allows to correct the Kohn-Sham eigenvalues, replacing the density-based exchange-correlation potential v^{XC} by the GW self-energy:

$$\varepsilon_n^{GW} = \varepsilon_n^{KS} + \langle\phi_n|\,\Sigma^{GW}(\varepsilon_n^{GW}) - v^{XC}\,|\phi_n\rangle$$

with ε_n^{GW} the so-called quasiparticle energies at which the self-energy operator must be calculated.

B. Fragment approximation

Our implementation[60,61] of the GW approach adopts a resolution-of-the-identity (RI) formalism[62–64] where the density and its variations are expressed over a Gaussian auxiliary basis set $\{P\}$. The auxiliary basis functions must thus approximate the space generated by the products of molecular orbitals (MO) $\{\phi_n\}$:

$$\phi_n(\mathbf{r})\,\phi_m(\mathbf{r}) \overset{RI}{\simeq} \sum_P \mathcal{F}_P(\phi_n\phi_m)\,P(\mathbf{r})$$

leading in particular to fitting coefficients $\mathcal{F}_P(\phi_n\phi_m) = 0$ when the ϕ_n and ϕ_m molecular orbitals are non-overlapping. Within this representation, Eqn. 4 rewrites:

$$\chi_0(\mathbf{r},\mathbf{r}';\omega) \overset{RI}{\simeq} \sum_{P,Q} X_0(P,Q\,;\omega)\,P(\mathbf{r})\,Q(\mathbf{r}') \tag{7}$$

with coefficients

$$X_0(P,Q\,;\omega) = \sum_{m,n}(f_m - f_n)\frac{\mathcal{F}_P(\phi_m^*\phi_n)\,\mathcal{F}_Q(\phi_m\phi_n^*)}{\omega - (\varepsilon_n - \varepsilon_m) + i\eta \times \mathrm{sgn}(\varepsilon_n - \varepsilon_m)}\,. \tag{8}$$

In the fragment approximation, the full system auxiliary basis is simply the union of the subsystem basis sets, while each subsystem density can be expressed in its own corresponding basis. In such a case, the analysis of Eqn. 8 indicates that the joint contribution of two non overlapping subsystems to the representation of the independent-electron susceptibility $\chi_0(\mathbf{r},\mathbf{r}';\omega)$ should be zero. Within that limit, the RI-fitted independent-electron susceptibility matrix is thus block diagonal, with blocks corresponding to the constituting subsystems gas phase (isolated) irreducible susceptibilities.

Labelling $\mathbf{X}_0^{(\mathrm{I})}$ the block of fit coefficients corresponding to the independent-electron susceptibility of subsystem (I), the Dyson equation for the full system interacting susceptibility coefficients matrix $\mathbf{X}(\omega)$ (eqn. 6) reads:

$$\mathbf{X}(\omega)^{-1} = \sum_{\mathrm{I}} \mathbf{X}_0^{(\mathrm{I})}(\omega)^{-1} - \sum_{\mathrm{I,J}} \mathbf{V}^{(\mathrm{IJ})} \tag{9}$$

with $\mathbf{V}^{(\mathrm{IJ})}$ the block corresponding to the Coulomb interactions between auxiliary basis elements of fragments (I) and (J). This equation can also be conveniently rewritten:

$$\mathbf{X}(\omega)^{-1} = \sum_{\mathrm{I}} \mathbf{X}_g^{(\mathrm{I})}(\omega)^{-1} - \sum_{\mathrm{I}\neq\mathrm{J}} \mathbf{V}^{(\mathrm{IJ})} \tag{10}$$

with $\mathbf{X}_g^{(\mathrm{I})}(\omega)$ the block of coefficients for the isolated (gas phase) interacting susceptibility of fragment (I). In this latter formulation, only the off-diagonal Coulomb interactions $\mathbf{V}^{(\mathrm{I}\neq\mathrm{J})}$ that accounts for inter-fragments coupling are considered (see Fig. 1). Similar equations can be found in the framework of subsystem TD-DFT.[65,66]

In the fragment approximation, the cost of calculating all fragments $\mathbf{X}_0^{(\mathrm{I})}$ and $\mathbf{X}_g^{(\mathrm{I})}$ blocks scales linearly with respect to the number of fragments. This is a considerable saving and it is now the inversion of the Dyson equation (Eqns. 3 or 10), with cubic scaling with respect to the total number of fragments, that becomes the bottleneck in the limit of a very large number of subsystems. Due to the $\mathbf{V}^{(\mathrm{IJ})}$ inter-blocks Coulomb matrix elements, the full interacting susceptibility $\mathbf{X}(\omega)$, and resulting screened Coulomb potential: $\mathbf{W}(\omega) = \mathbf{V} + \mathbf{V}\mathbf{X}(\omega)\mathbf{V}$, are no longer diagonal by blocks. As such, the inversion of the Dyson equation for $\mathbf{X}(\omega)$, rewriting Eq. 10 as:

$$\mathbf{X}(\omega) = \mathbf{X}_g(\omega) + \mathbf{X}_g(\omega)\mathbf{V}^{\mathrm{off}}\mathbf{X}(\omega), \tag{11}$$

where $\mathbf{X}_g(\omega) = \sum_{\mathrm{I}} \mathbf{X}_g^{(\mathrm{I})}(\omega)$ and $\mathbf{V}^{\mathrm{off}} = \sum_{\mathrm{I}\neq\mathrm{J}} \mathbf{V}^{(\mathrm{IJ})}$, becomes a huge effort in terms of computation time and memory in the limit of large systems. It is such a problem that we address below by optimally reducing the size of each $\mathbf{X}_g^{(\mathrm{I})}(\omega)$ susceptibility representation.

C. Constrained reduction of the fragment susceptibilities

A typical calculation that expands MOs over a triple-zeta def2-TZVP basis set[67] involves the corresponding optimized auxiliary def2-TZVP-RI basis set[68] that is composed of 95 orbitals for e.g. B, C, N, O atoms. In this situation, an environment containing of the order of 10^5 atoms will result in susceptibility matrices of the order of 10^7 in size to be dealt with in the Dyson equation.

For the fragment of interest, for which we want to actually perform a *GW* correction, we preserve the full auxiliary basis optimized for the corresponding MO basis sets. However, concerning the fragments in the environment, we emphasize that we are mainly interested in their contribution to the reaction field, namely to the induced dipoles, quadrupoles, etc., developed as a response to an electronic excitation in the central subsystem. As such, the full details of the susceptibility in the auxiliary $\{P\}$ basis may not be necessary.

In order to reduce the computational effort of the Dyson equation, we therefore look for an efficient and compact way to represent the interacting susceptibility of the fragments in the environment. We seek a lower-rank approximation to the gas phase interacting susceptibilities $\chi_g^{(\mathrm{I})}(\omega)$:

$$\chi_g^{(\mathrm{I})}(\mathbf{r},\mathbf{r}';\omega) \overset{RI}{\simeq} \sum_{P,Q} X_g^{(\mathrm{I})}(P,Q\,;\omega)\; P(\mathbf{r})\, Q(\mathbf{r}') \tag{12}$$

$$\overset{MODEL}{\simeq} \sum_{\gamma,\gamma'} \widetilde{X}_g^{(\mathrm{I})}(\gamma,\gamma'\,;\omega)\; \gamma(\mathbf{r})\, \gamma'(\mathbf{r}') \tag{13}$$

where we use a small basis sets $\{\gamma\}$, which could be for example a minimal Gaussian (sp^3) 4-orbitals basis per atom in order to mimic the induced charges-and-dipoles models developed in QM/MM techniques.[10]

For the isolated fragment (I), the resulting errors in the interacting susceptibility and its corresponding contribution to the screening, or reaction, field respectively read:

$$\begin{aligned}\Delta\chi_g^{(\mathrm{I})}(\mathbf{r},\mathbf{r}';\omega) = &\sum_{\gamma,\gamma'} \widetilde{X}_g^{(\mathrm{I})}(\gamma,\gamma'\,;\omega)\; \gamma(\mathbf{r})\, \gamma'(\mathbf{r}') \\ &- \sum_{P,Q} X_g^{(\mathrm{I})}(P,Q\,;\omega)\; P(\mathbf{r})\, Q(\mathbf{r}')\end{aligned} \tag{14}$$

and

$$\begin{aligned} &\Delta v^{(\mathrm{I})}_{\mathrm{screen}}(\mathbf{r},\mathbf{r}';\omega) \\ &= \iint \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2\;\; v(\mathbf{r},\mathbf{r}_1)\,\Delta\chi^{(\mathrm{I})}_g(\mathbf{r}_1,\mathbf{r}_2;\omega)\,v(\mathbf{r}_2,\mathbf{r}'). \end{aligned} \tag{15}$$

Once the model "polarization" basis $\{\gamma\}$ is fixed, the associated $\widetilde{\mathbf{X}}^{(\mathrm{I})}_g$ set of coefficients can thus be simply deduced by minimizing the error in the screening field $\Delta v^{(\mathrm{I})}_{\mathrm{screen}}$ through a set $\{t\}$ of test functions:

$$\widetilde{\mathbf{X}}^{(\mathrm{I})}_g(\omega) = \underset{\left\{\widetilde{X}^{(\mathrm{I})}_g(\gamma,\gamma';\omega)\right\}}{\mathrm{argmin}} \sum_{t,t'} \left|\langle t\,|\Delta v^{(\mathrm{I})}_{\mathrm{screen}}(\omega)|\,t'\rangle\right|^2. \tag{16}$$

The salient features of this equation is that, since we measure the difference in the reaction field on a set of test functions, we are free to focus on the specific components of the screening field $v\chi v$ that we want to preserve. Details about the resolution of this equation are given in Appendix A.

A naive choice for the test functions $\{t\}$ could be the set of auxiliary functions $\{P\}$ used to construct the reference $\chi^{(\mathrm{I})}_g$ response function. As shown below, this strategy is rather inefficient. The reason for this failure is that we do not intend to use the model susceptibility to perform a *GW* calculation on fragment (I) itself. Instead, we need it to build the fully interacting screening potential $v\chi v$ through the Dyson equation (Eq. 10) that couples (I ≠ J) fragments, focusing in the end on the central fragment (I = 0 in Fig. 2) on which we perform the *GW* calculation. As such, priority should be given to the interactions between the different model fragments, starting from neighboring fragments up to the long range interactions dominated by low order momenta of their polarizability tensors.

A very simple yet successful strategy consists in keeping the test functions localized on the atoms of the fragment (I) for which we seek the model susceptibility, but building the test set $\{t\}$ out of very diffuse orbitals that will sample the surrounding fragments. The basic idea is that using such diffuse test functions allows to "reach out" for the effect of $v\chi^{(\mathrm{I})}v$ on neighboring molecules. The test set can then be completed with the auxiliary basis $\{P\}$ associated with fragment (I), but down-weighted, in order to keep the emphasis on the diffuse orbitals during the minimization process. More details about this test basis can be found in Appendix B.

Simultaneously, the preservation of long range interactions can be guaranteed by enforcing low order Cartesian momenta of the susceptibility through Lagrange multipliers:

$$\left\langle x^m\,y^n\,z^p\;\middle|\;\Delta\chi^{(\mathrm{I})}_g(\omega)\middle|\;x^{m'}\,y^{n'}\,z^{p'}\right\rangle = 0. \tag{17}$$

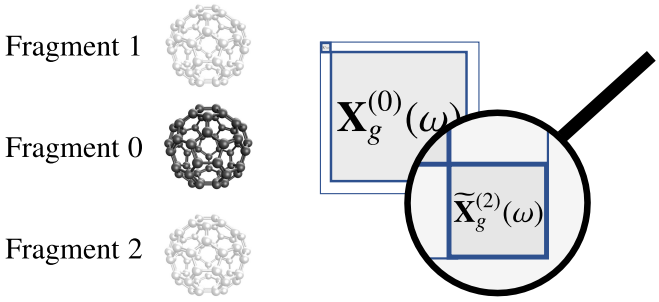


FIG. 2. Sketch of the susceptibility blocks compression associated with the fragments in the environment (shaded C_{60}). The susceptibility of the central fragment of interest (fragment I=0), on which will be performed the embedded *GW* calculation, is not reduced.

In the following, we label l_{max} the maximum order $m + n + p$ enforced for a specific model susceptibility. For example, $l_{max} = 1$ corresponds to the preservation of the fragment neutral monopole, as well as the dipolar polarizability tensor. Imposing such a constraint, along with the use of diffuse test functions, ensures that the reaction field will be well reproduced not only in the vicinity of fragment (I), but in the long-range as well. Imposing higher order polarizability tensors momenta (e.g. $l_{max} = 2$), can be achieved but with a mild impact as discussed below.

D. Minimal effective polarizability basis

So far the choice of the $\{\gamma\}$ has been left arbitrary. Contrary to the induced charge-and-dipole models used in semi-empirical QM/MM techniques, we follow here a more automated route. The polarization basis can be obtained as the result of a generalized minimization process, that is we include the $\{\gamma\}$ in the minimization process:

$$\underset{\{\gamma\}}{\text{argmin}} \left(\min_{\left\{\widetilde{X}_g^{(\mathrm{I})}(\gamma,\gamma';\omega)\right\}} \sum_{t,t'} \left| \langle t \,|\Delta v_{\text{screen}}^{(\mathrm{I})}(\omega)|\, t' \rangle \right|^2 \right) \tag{18}$$

where the only input choice is now the number N_P of γ polarization vectors. In practice, the γ functions are expressed in the $\{P\}$ auxiliary basis set associated with fragment (I), namely

$$\gamma(\mathbf{r}) = \sum_P C_{\gamma P}\, P(\mathbf{r}) \,, \tag{19}$$

so that the $\{C_{\gamma P}\}$ coefficients are now the minimization variables. Formally, setting N_p to the dimension of the original auxiliary basis leads to a model susceptibility $\widetilde{\mathbf{X}}_g$ strictly equal to the reference $\mathbf{X}_g$. Such an equality holds independently of the choice of the test basis or the constraints.

We tackle this somewhat complex minimization problem by iterating over two distinct steps: i) inner optimization of the $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}(\omega)$ matrix elements at fixed $\{\gamma\}$ (see equation 16), which is solved exactly through linear algebra; ii) outer optimization of the $\{C_{\gamma P}\}$ coefficients at fixed $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}(\omega)$ matrix elements which is done using gradient descent techniques. More details about this last step can be found in Appendix C.

In the simplest case where test functions $\{t\}$ only span the auxiliary $\{P\}$ set, and in the absence of constraint, the Eckart–Young–Mirsky theorem states that the N_P optimal $\{\gamma\}$ polarization vectors are similar to those defined in Refs. 69,70, namely the leading eigenvectors of the so-called symmetrized susceptibility. The present minimization formulation allows further flexibility with the introduction of test functions and constraints, allowing on-the-fly design of model dielectric functions for specific purposes, emphasizing short-to-long-range or on-site accuracy.

We conclude this Section by emphasizing again that the operations described above (calculations of the reference $\mathbf{X}_g^{(\mathrm{I})}(\omega)$ and $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}(\omega)$ matrices, SVD decomposition of related operators, etc.) are performed on isolated fragments, leading to a computational cost that is linear in the number of distinct fragments. In turn, the number of operations related to inverting the Dyson equation for the total screened Coulomb potential, involving interactions between all fragments, is dramatically reduced through reduction of the associated prefactor. As shown below, the optimal polarization basis can be made typically 10^2 times smaller than the original auxiliary basis set, preserving the polarization energy in the meV range, leading to a reduction of the order of 10^6 of the cost associated with obtaining an accurate W operator on the central fragment (I=0).

E. Technical details

The present subsystem approach with minimal representation of the fragments electronic susceptibility has been implemented in the BEDEFT (beyondDFT) package.[60,61] Input Kohn-Sham eigenstates are generated at the def2-TZVP PBE0[71,72] level with the ORCA package.[73,74] We adopt the corresponding def2-TZVP-RI auxiliary basis sets associated with the Coulomb-fitting resolution-of-the-identity (RI-V) approach.[62,75] The molecular geometries for C_{60} and the pentacene are obtained at the def2-TZVP PBE0 level. The face-centered cubic (fcc) C_{60} dense phase is constructed taking experimental lattice parameters[76] (a = 14.17 Å), neglecting orientational disorder. The C_{60} surface we consider is the (111) surface.

Even though the present scheme allows to compute reaction fields at finite (imaginary) frequen-

cies, following the analytic continuation approach to the dynamical GW self-energy implemented in BEDEFT,[60] we calculate here the polarization energies at the static Coulomb-Hole plus Screened-Exchange (COHSEX) level,[12] with:

$$\Sigma^{\mathrm{SEX}}(\mathbf{r},\mathbf{r}') = -\sum_{n}^{\mathrm{occp}} \phi_n(\mathbf{r})\,\phi_n^*(\mathbf{r}')\,W(\mathbf{r},\mathbf{r}';\omega=0) \tag{20}$$

$$\Sigma^{\mathrm{COH}}(\mathbf{r},\mathbf{r}') = \frac{1}{2}\,\delta(\mathbf{r}-\mathbf{r}')\left[W(\mathbf{r},\mathbf{r}';\omega=0) - v(\mathbf{r},\mathbf{r}')\right] \tag{21}$$

with the screened-exchange term involving a summation over occupied (occp) levels only. Following previous studies,[26,27,41–44] our polarization energy P_n for a given energy level is taken to be the difference between the static COHSEX energy level in the presence of a polarizable environment and its analog in the gas phase, namely:

$$P_n = \varepsilon_n^{GW_{\mathrm{e}}} - \varepsilon_n^{GW_{\mathrm{g}}} \simeq \underbrace{\varepsilon_n^{\mathrm{COHSEX}_{\mathrm{e}}} - \varepsilon_n^{\mathrm{COHSEX}_{\mathrm{g}}}}_{=P_n^{\Delta\mathrm{COHSEX}}}$$

where the index (e) and (g) in $GW_{\mathrm{e/g}}$ and $\mathrm{COHSEX}_{\mathrm{e/g}}$ stand for embedded (e) and gas (g) phases. The static COHSEX formalism can be understood as an approximation to the GW operator in the limit where the dielectric response is taken to be energy-independent, as indicated by the ($\omega = 0$) restriction in Eqs. 20 and 21. Since only the static susceptibility needs to be calculated, it is one order of magnitude cheaper than a full GW approach requiring the calculation of $n\omega \simeq 12$ independent-electron $\chi_0(i\omega)$ susceptibilities along the imaginary axis in our analytic continuation scheme. The accuracy of using such a static limit for the polarisation energies induced by a neighboring dielectric medium was first analyzed in the case of molecules close to a metallic substrate.[47] Such an approximation, that considers only the low-frequency limit of the environment electronic dielectric response, is consistent with standard PCM implementations where the macroscopic dielectric constant originating from electronic degrees of freedom is taken to be the optical one in the low frequency limit (square of the refractive index). Similarly, in standard QM/MM implementations, the semi-empirical atomic polarizabilities are designed to reproduce the fragment electronic polarizability in the static limit.[10] Extension to dynamical reaction fields will be discussed in subsequent studies.

From the knowledge of such a polarization energy P_n, calculated at the static ΔCOHSEX level, the absolute quasiparticle energy can be obtained as:

$$\varepsilon_n^{GW_{\mathrm{e}}} = \varepsilon_n^{GW_{\mathrm{g}}} + P_n^{\Delta\mathrm{COHSEX}}.$$

In the fragment approximation, and in the absence of wavefunction hybridization, such a value yields the energy of the corresponding band center.[77] In particular, one can recover the experimental peak-to-peak gap in the dense phase, namely the difference of energy between the highest-occupied and lowest-unoccupied molecular orbital (HOMO/LUMO) band centers. In the following, when needed, the gas phase $\varepsilon_n^{GW_g}$ quasiparticle energy levels will be calculated at the partially self-consistent evGW@PBE0 level, that has been shown[78,79] to be more accurate than non-self-consistent calculations, unless an optimally tuned functional is used for the starting DFT Kohn-Sham calculation.[78,80]

Finally, we only need to perform an explicit GW correction for the fragment of interest (I = 0). In other words, while the interacting susceptibility matrix $\mathbf{X}(\omega = 0)$ of Eq. 10 is defined for the full system, the screened Coulomb potential matrix $\mathbf{W} = \mathbf{V} + \mathbf{VXV}$ entering Eqs. 20 and 21 is only computed explicitly for the corrected fragment. This enables us to save both on memory footprint and CPU time aspects.

III. RESULTS

A. Validation

We start by looking at the evolution of the static dipolar polarizability tensor for a given fullerene, in the gas phase, obtained with the model susceptibility matrix as a function of N_P (the number of polarization vectors we keep). Namely, we compute

$$[\alpha]_{ij} = -\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\, r_i \left(\sum_{\gamma,\gamma'}^{N_P} \gamma(\mathbf{r}) \widetilde{X}_g(\gamma,\gamma';\omega=0)\gamma'(\mathbf{r}') \right) r'_j. \tag{22}$$

This tensor is a key quantity for the long-range screening effects originating from a given fragment and represents thus a direct measure of the accuracy of the fitted susceptibility. First, we explore the strategy where the test functions $\{t\}$ are taken to span the auxiliary basis $\{P\}$ located on the fragment (a fullerene) for which we build the model susceptibility. Relative errors on the (Frobenius) norm of the dipolar polarizability tensor, with respect to a reference calculation using the full auxiliary basis (5700 vectors), are represented in Fig. 3 (blue dots). As expected, this error decreases as the number of polarization vectors increases. For $N_P = 240$, the relative error is of the order of 10%, that is still rather large. This number $N_P = 240$ corresponds to a typical minimal sp^3 basis per atom of the kind used in semi-empirical induced charges-and-dipoles polarizable models.

We now perform the same exercise but adding to the test functions $\{t\}$ a set of atom-centered

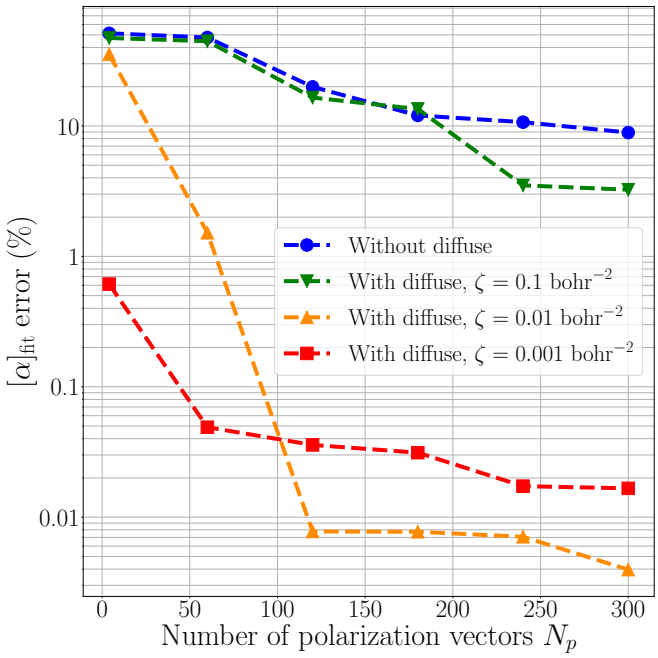


FIG. 3. Evolution of the relative error $\left\|[\alpha]_{\text{fit}} - [\alpha]_{\text{ref}}\right\| / \left\|[\alpha]_{\text{ref}}\right\|$ (in percentage) of a fullerene RPA dipolar polarizability tensor $[\alpha]$ (see Eq. 22) as a function of the number N_P of polarization vectors (maximum number 5700). $\|\cdot\|$ corresponds to the Frobenius norm. $[\alpha]_{\text{ref}}$ is computed with the susceptibility described in the full auxiliary basis (with 5700 vectors), while $[\alpha]_{\text{fit}}$ is calculated using only N_p polarization vectors. Results for different choices of test functions are plotted.

diffuse Gaussian orbitals. Such diffuse functions are typically one set of (*s,p,d,f,g*) orbitals per atom with, for sake of simplicity, the same $e^{-\zeta r^2}$ radial part. Results for different values of ζ are reported on Fig. 3. A value of $\zeta = 0.1$ bohr^{-2} (green down triangles), comparable to 0.2 bohr^{-2} for the most diffuse carbon atomic orbital in the def2-TZVP-RI basis set, does not improve the quality of the fit. Increasing the diffuse character of these functions, with $\zeta = 0.01$ bohr^{-2} (orange up triangles) improves significantly the quality of the results. For $N_p = 120$, namely $\simeq 2\%$ of the dimension of the original auxiliary $\{P\}$ basis set, the relative error is below 0.01%. Increasing too much the extent of the diffuse orbitals, with e.g. $\zeta = 0.001$ bohr^{-2} (red squares) degrades the quality of the results. Even if relative errors are smaller with such a small ζ value in the limit of a very small number of polarization vectors ($N_p = 4$ or 60), $\zeta = 0.001$ bohr^{-2} leads to greater errors than $\zeta = 0.01$ bohr^{-2} for larger values of N_p.

The quality of the polarizability tensor obtained with the low-rank susceptibility insures that long-range interactions will be accurately reproduced. We now focus on nearest-neighbor interactions.

We study in particular the HOMO/LUMO energy gap associated with a fullerene (in red in Fig. 4 Inset) surrounded by its first shell of 12 nearest-neighbors (in blue).

In a standard fragment calculation at the full def2-TZVP/def2-TZVP-RI level, the central C_{60} HOMO-LUMO gap closes by ~0.98 eV due to the enhanced screening induced by the first shell of neighbors. This represents about 60% of the total polarization energy (see Section III B) as compared to a fullerene in a fullerite, namely an infinite fullerene crystal.

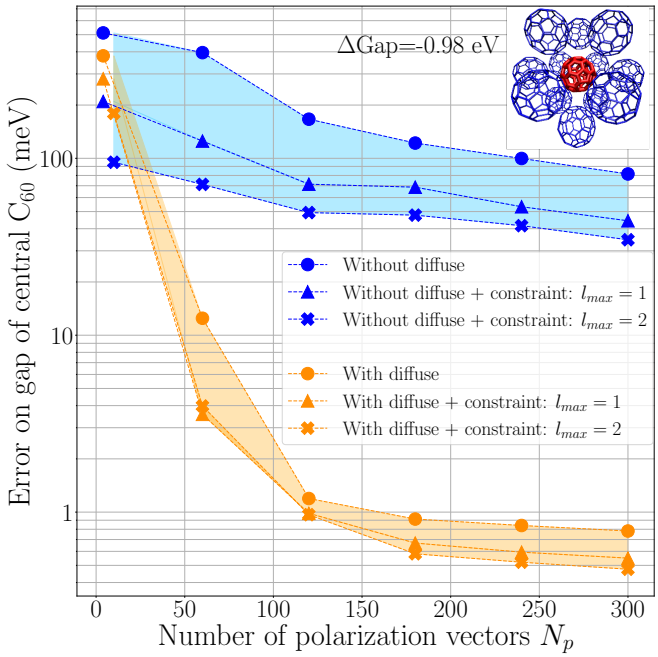


FIG. 4. Error on the central C_{60} [red, Inset] gap as a function of the number N_P of polarization vectors per C_{60} in the first-neighbors shell [blue, Inset]. Energies on the ordinates are in meV and log-scale. Results for test functions with and without diffuse orbitals, with and without constraints (see section II C), are shown. Diffuse functions use $\zeta = 0.01$ bohr^{-2}. The reference ΔGap $= -0.98$ eV corresponds to the gap reduction from the gas phase to the 13-C_{60} cluster when all fragments are described by their full basis ($N_p = 5700$).

We now study the effect of reducing the size N_P of the polarization basis on the 12 surrounding C_{60}. As previously, we start by using test functions taken only in the span of the auxiliary basis $\{P\}$ of the fragment (a fullerene) whose model susceptibility is fitted. The results are provided on Fig. 4 (blue dots). As expected, the error on the central C_{60} HOMO-LUMO gap, as compared to the reference calculation, decreases with the number of polarization vectors. For $N_P = 240$, the error is of the order of 100 meV, allowing to have a qualitative result but representing still an error of the order of 10% with respect to the targeted polarization energy. As discussed above, this error

converges to zero with N_p equal to the size of the original atom-centered Gaussian auxiliary basis set (5700 vectors), but the convergence is too slow.

Similarly to the previous study of the polarizability, we now add diffuse functions in the test basis, with $\zeta = 0.01$ bohr^{-2}. The related evolution of the error on the polarization energy for the gap is represented in Fig. 4 (orange dots). Clearly, the addition of diffuse functions, allowing to test the quality of the model $\mathbf{V}\widetilde{\mathbf{X}}_g^{(\mathrm{I})}\mathbf{V}$ reaction field in the vicinity of molecule (I), dramatically accelerates the convergence of the polarization energy with respect to the size of the model susceptibility matrix. For $N_P = 180$, namely $\simeq 3\%$ of the original susceptibility matrix size (5700) for one C_{60}, the error is now of the order of 1 meV, reaching quantitative accuracy.

We further add the constraints (equation 17), with and without diffuse functions, to enforce the exact dipolar polarizability tensor with $l_{max} = 1$ (up triangles in Fig. 4), or up to second order moments with $l_{max} = 2$ (crosses in Fig. 4). In all cases, the constraints improve the accuracy, in particular in the small N_P limit, even though their impact is not as important as adding diffuse test functions. Such a behaviour can be understood by looking, e.g., at Fig. 3 for $\zeta = 0.01$ bohr^{-2} and $N_P = 120$. The dipolar polarizability is already quite well reproduced so that the constraint leads to a small improvement. Fig. 4 reveals that the constraint $l_{max} = 2$ improves very slightly the error on the gap, in comparison to the constraint $l_{max} = 1$. When diffuse function are added to the test basis, the differences between $l_{max} = 1$ (orange up triangles) and $l_{max} = 2$ (oranges crosses) are less than 0.1 meV for $N_p \geq 120$. Since imposing the constraint comes at no cost, we keep $l_{max} = 1$ in the forthcoming calculations.

We provide in the Supplementary Material (Fig. S1) the same study but for the HOMO and LUMO energies. Very similar results are obtained as compared to the gap study. In particular, including diffuse orbitals in the test set, together with the $l_{max} = 1, 2$ constraints, the errors on the HOMO and LUMO are of the order of 3 to 4 meV, namely just slightly larger than the sub-meV error obtained on the gap.

Very similar results can also be obtained by considering Kohn-Sham basis sets, and associated auxiliary basis sets, containing diffuse orbitals. As a matter of fact, it is well documented that the description of response properties, such as the polarizability, converges faster when diffuse channels are included for the description of molecular orbitals and associated density.[81,82] For sake of consistence with our def2-TZVP/def2-TZVP-RI calculations, we adopt the def2-TZVPD/def2-TZVPD basis sets[82] that add diffuse channels for better response properties. The resulting data are presented in the Supplementary Material (Fig. S2). Due to the addition of diffuse orbitals in

the Kohn-Sham basis, the closing of the gap increases by about 2%, from ΔGap= -0.98 eV to ΔGap= -1.00 eV, indicating a slightly enhanced screening, consistent with an enhanced molecular polarizability. The evolution of the error with the number N_p of polarization vectors is also very similar to what we obtained in Fig. 4, demonstrating that our scheme (diffuse test functions and constraints) is universal, that is rather independent of the chosen input Kohn-Sham basis set.

The test provided above for the polarization energy originating from the first-nearest-neighbors is the most stringent test. For fragments located farther away, the dipolar component of the reaction field, that we strictly impose, becomes more and more dominant. This is illustrated in Fig. 5 where we study the HOMO-LUMO gap of a C_{60} surrounded now by its two nearest-neighbor shells (see Inset Fig. 5). The size of this cluster amounts to 55 fullerenes. When all fullerenes are described at their full def2-TZVP/def2-TZVP-RI level (in the fragment approximation), the gap of the central fullerene closes by 1.25 eV from the gas phase to the 55-C_{60} cluster geometry. We study the effect of reducing the size N_p of the polarization basis used to describe the susceptibility of the 42 surrounding C_{60} in the second shell, keeping the full auxiliary basis to describe the central fullerene and its first-nearest neighbors shell. As such, we mainly focus on the error induced by the fitting process on fragments located at middle to long-range of the central subsystem of interest.

Consistently with the results obtained for the first shell of neighbors (Fig. 4), these calculations confirm that the addition of diffuse orbitals in the test set dramatically helps in reducing the error below the meV with a small number of N_P polarization vectors (compare orange and blue data in Fig. 5). Further, as compared to the first-nearest neighbors case, a smaller number N_P of polarization vectors is needed to go below the meV error when the constraint on the dipolar polarizability ($l_{max} = 1$) is imposed. This is the signature that in the long-range, the dipolar response dominates the screening potential, or reaction field. The combination of diffuse test orbitals with $\zeta = 0.01$ bohr^{-2} with the constraint $l_{max} = 1$ leads to an error of the order of 0.1 meV for $N_P = 60$, namely one polarization vector per atom. This is a dramatic reduction of the size of the polarization basis needed to describe the susceptibility blocks entering the Dyson equation.

As an additional validation, we plot in Fig. 6 the screening potential, or reaction field, $v_{\text{screen}}(\mathbf{r}, \mathbf{r}_0) = [v\chi v](\mathbf{r}, \mathbf{r}_0)$ associated with an elementary positive source point-charge located in $\mathbf{r}_0$, in the vicinity of a single pentacene molecule. Namely, for a test charge located in $\mathbf{r}_0$, we plot $v_{\text{screen}}(\mathbf{r}, \mathbf{r}_0)$ as a function of $\mathbf{r}$ in the pentacene plane. In the case of a single fragment molecule, the reaction field reduces to $v\chi^{(\text{I})}v$, with (I) the index of that molecule. The reference reaction field is provided in Fig. 6(a) while the error associated with the model susceptibility, for a fixed $N_P = 70$

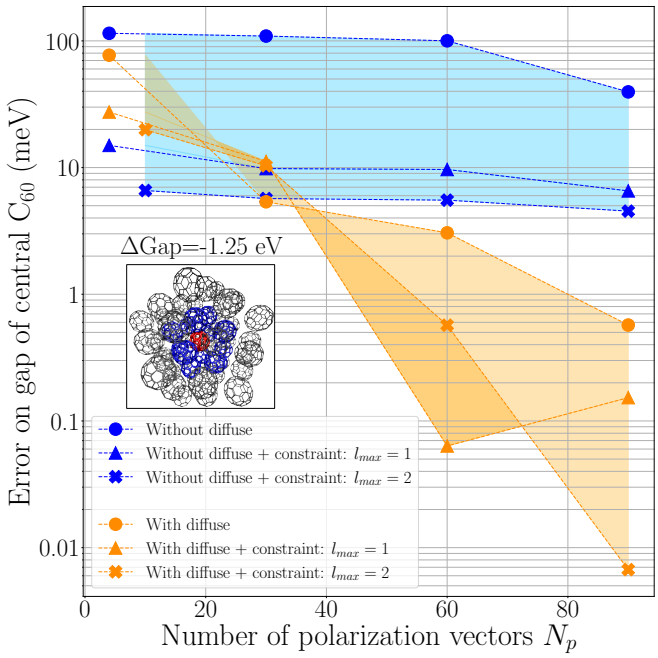


FIG. 5. Error on the gap for a fullerene surrounded by its two first-shells of neighbors. The susceptibility of the central (in red, Inset) and 12 first-nearest-neighbors (in blue, Inset) C_{60} are described by the full auxiliary basis (5700 orbitals), while the susceptibility for each of the 42 C_{60} in the second shell of neighbors (in grey, Inset) is described by N_P polarization vectors. Energies on the ordinates are in meV and log-scale. Results for test functions with and without diffuse orbitals, with and without constraints (see section II C), are shown. Diffuse functions use $\zeta = 0.01$ bohr^{-2} (see text). The value $\Delta\text{Gap} = -1.25$ eV corresponds to the reference gap reduction for the central C_{60} from the gas phase to this 55-fullerenes cluster configuration when all fragments are described with their full basis sets.

number of retained polarization vectors among the 2314 vectors of the original def2-TZVP-RI basis, is represented in the other subfigures. Fig. 6(b) shows the case where the test basis set does not contain diffuse functions. In Fig. 6(c) we add diffuse functions ($\zeta = 0.01$ bohr^{-2}), while Fig. 6(d) illustrates the fit of χ with the same diffuse functions and the constraint $l_{max} = 1$. Clearly, the error associated with the reaction field around a given pentacene molecule (in meV units) is dramatically reduced upon adding diffuse test functions and the constraint on the dipolar polarizability. We purposely replaced the fullerene molecule by a pentacene to indicate that the accuracy of the present scheme is hardly system dependent.

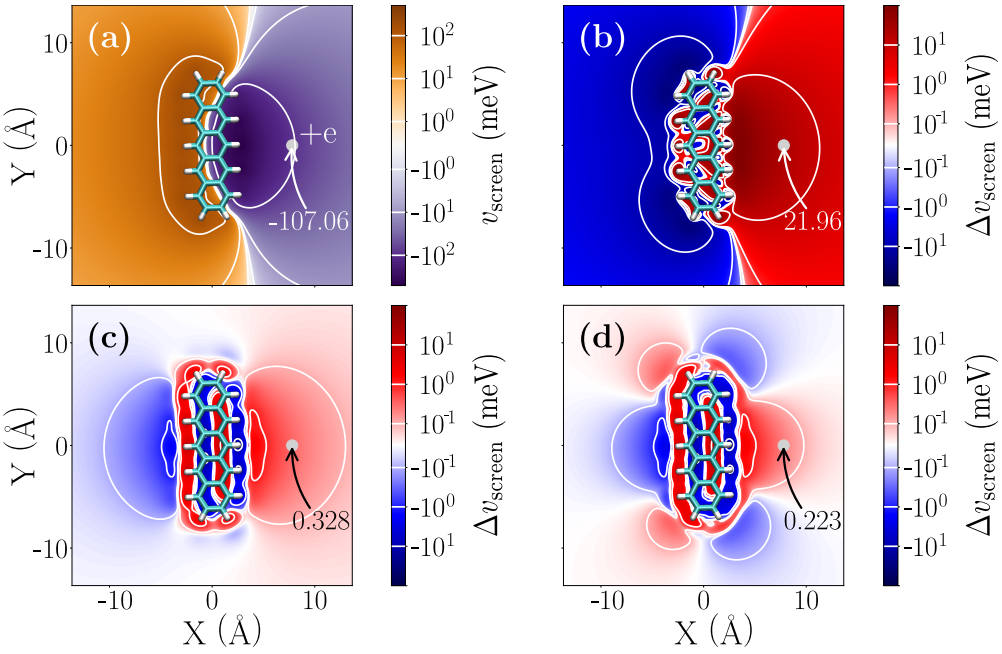


FIG. 6. In plane $[v\chi v](\mathbf{r}, \mathbf{r}_0; \omega = 0)$ static reaction field generated by a pentacene molecule in response to a positive unit source charge (grey dot indicated by +e) in $\mathbf{r}_0$. In (a), the full reaction potential. In (b), (c) and (d) the error $\Delta v_{\rm screen}(\mathbf{r}, \mathbf{r}_0; \omega = 0)$ (see Eq. 15) with respect to the full reaction potential upon substituting the full χ by its low-rank approximation (with 70 polarization vectors). In (b), low-rank χ model obtained without diffuse functions in the $\{t\}$ test set; in (c) adding diffuse orbitals to the $\{t\}$ test set, and in (d) adding further the constraint on the dipolar molecular polarizability. Positions are in angstroms, and units of the reaction field are in meV and in log scale. Equipotentials at $(\pm 1, \pm 10, \pm 100)$ meV in figure (a), and at $(\pm 0.1, \pm 1, \pm 10)$ meV in figures (b), (c) and (d) are represented by white lines. The value of the reference and of the error associated with the reaction field at the position of the source is indicated.

B. The C_{60} crystal and surface environments

Beyond the small cluster models, we now study the evolution of the gap of a fullerene from the gas phase to a C_{60} face-centered-cubic crystal (fcc) environment. We thus want to calculate the closing of the gap by screening effects in the limit of an infinite environment. Such a quantity, labeled ΔGap below, is also coined the polarization energy. We will consider the cases of bulk C_{60} (fullerite) and further of a C_{60} at the (111) surface and sub-surface. Experimental photoemission experiments are very much surface sensitive for organic systems, with a limited penetration depth of the input photons or electrons, so that comparison with the surface location is more appropriate. The bulk limit is obtained by immersing a C_{60} molecule in a sphere of fullerenes with increasing radius.

The surface and subsurface limits are obtained by using half-a-sphere of polarizable fullerenes in the environment. In the absence of wavefunction delocalization (or band dispersion) in the fragment approach, we focus on the peak-to-peak gap, namely the gap between the center of the HOMO and LUMO bands. We emphasize that the absence of permanent ground-state dipole, quadrupole, etc., in fullerene molecules, precludes the influence of any electrostatic crystal field in the ground-state.

In the present case of fullerene crystals, the reference $\mathbf{X}_g^{(\mathrm{I})}$ susceptibility can be constructed for a single fullerene and the resulting fitted $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}$ matrix can be "copied" to form the model susceptibility block associated with each fullerene in the environment. Even though rotational disorder was not explored in this study, rotating the $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}$ matrix, to follow the rotation of a given fullerene, can be easily implemented. Beyond rotations, the effect on the polarization energy of changes in the susceptibility matrix associated with slight atomic distortions around some average equilibrium geometry, is expected to be small but may be explored in future studies.

The calculations are performed with the parameters described above, namely keeping the full auxiliary def2-TZVP-RI basis set to describe the susceptibility of the C_{60} of interest for which we perform our embedded *GW* calculation. The same full auxiliary basis set is used for its first-nearest neighbors. For the rest of the environment, namely the second-nearest neighbors and those located farther away, we keep $N_P = 60$ optimized polarization vectors for each fullerene. Diffuse test functions with $\zeta = 0.01$ bohr^{-2} are adopted with the $l_{max} = 1$ constraint. We focus on the gap closing (ΔGap) from the gas to the dense phase. Such a polarization energy, originating from the screening by the environment, is described at the ΔCOHSEX level as emphasized above.

We plot in Fig. 7 the evolution of the polarization energy, or ΔGap, associated with the peak-to-peak gap for a bulk C_{60} (orange dots), a surface C_{60} (blue dots) and a C_{60} at sub-surface (red dots), as a function of $1/N_{C_{60}}^{1/3}$, where $N_{C_{60}}$ represents the number of fullerenes retained in the sphere or half-sphere we use for the environment. Analytic derivations show indeed that such a polarization energy converges slowly with a $1/N_{C_{60}}^{1/3}$ behavior in the asymptotic regime. This asymptotic behavior is confirmed numerically in Fig. 7 by the straight dashed-line fits, one for each type of system, going through the calculated energies in the large $N_{C_{60}}$ limit.

In the asymptotic infinite bulk size limit, the gap of the central C_{60} is closing by 1.69 eV (orange dashed line) with respect to the gas phase. This can be compared to the polarization energy of the biggest studied system. Our calculations are performed for spheres containing up to 3367 C_{60}, representing 202 020 carbon atoms. The gap of such a system closes by 1.58 eV, which represents a difference of 0.11 eV with respect to the extrapolated infinite size value. This highlights the

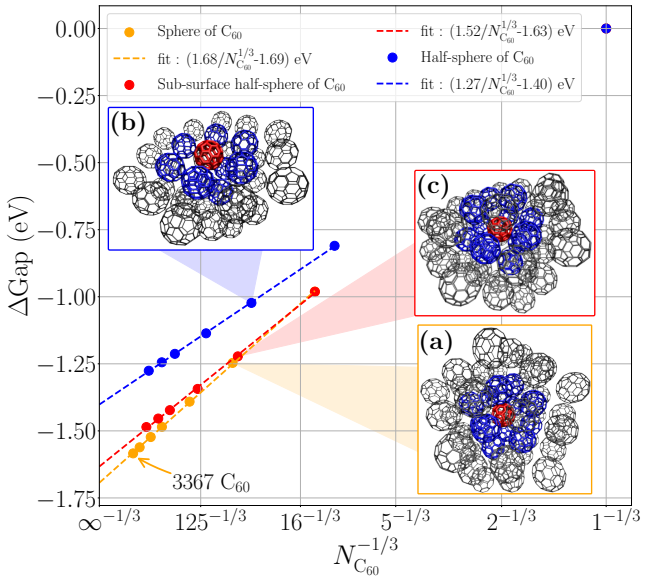


FIG. 7. Evolution of the gap closing ΔGap, at the static ΔCOHSEX level, between a gas phase fullerene and $N_{C_{60}}$ systems, as a function of $1/N_{C_{60}}^{1/3}$. The orange/blue/red dots show calculations for one C_{60} embedded at the center of a sphere [Inset (a)], at the surface of a half-sphere [Inset (b)], and at the subsurface of a half-sphere [Inset (c)], respectively. Dashed lines represent $[a/N_{C_{60}}^{1/3} + b]$ fits.

difficulty to capture the polarization energy with an accuracy within the 0.1 eV threshold when limited size environments are considered.

In order to compare to the G_0W_0@LDA 3.0 eV peak-to-peak gap in a fully periodic bulk calculation performed by Shirley and Louie in their pioneering study,[83] we compute the def2-TZVP G_0W_0@LDA gap for an isolated fullerene. Subtracting the 1.69 eV polarization energy in the bulk limit to our gas phase 4.46 eV G_0W_0@LDA HOMO-LUMO gap, we end up with a 2.8 eV peak-to-peak gap, in fair agreement with the periodic G_0W_0@LDA value.

We also compute the asymptotic infinite limit for the gap of one C_{60} located at the (111) surface (see Inset Fig. 7(b)). The blue dashed line gives an asymptotic closing of the gap amounting to 1.40 eV, which is in good agreement with the experimental values of 1.1 eV,[84] 1.2 eV[85,86] or 1.4 eV.[87,88] Such values were obtained by taking the provided experimental peak-to-peak gaps subtracted to the experimental 4.9 eV gap value for a fullerene in gas phase.[89] Alternatively, using the analytic continuation approach to the dynamical *GW* self-energy implemented in BEDEFT,[60] we compute the def2-TZVP ev*GW*@PBE0 gap value for one C_{60} in the gas phase. In the partially self-consistent ev*GW* approach, the HOMO and LUMO energy levels are converged at the 0.1

meV level at the 7th cycle. Taking the computed 5.1 eV gas phase gap value, and adding the calculated polarization energy at the surface, one obtains a surface peak-to-peak gap of 3.7 eV, within the 3.5-3.8 eV experimental range. The presence of a metallic substrate when performing photo-emission experiments, potentially enhancing the screening in the limit of few C_{60} layers, and alternatively the limited screening from the fullerene crystal in the few layer limits, may explain variations between experimental values. On the theoretical side, the influence of the fragment approximation, together with treating screening effects in the static COHSEX limit, remains to be studied.

We further compute the polarization energy for one C_{60} at the sub-surface (see Inset Fig. 7(c)). The red dashed fit provides an asymptotic infinite size polarization energy of 1.63 eV. This value, closer to the bulk limit (difference of 0.06 eV) than the surface limit (difference of 0.23 eV), tends to show a rapid convergence of the polarization energy with respect to the depth of the considered fullerene. Namely, a fullerene in the subsurface presents properties already close to bulk case.

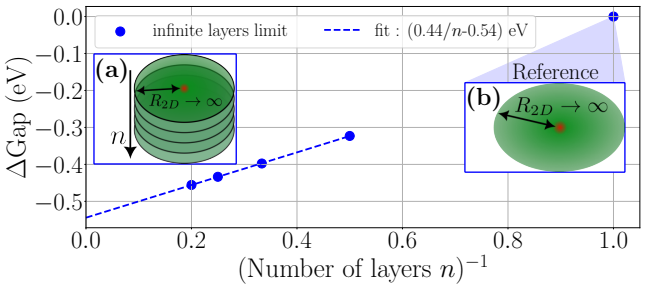


FIG. 8. Gap evolution for one surface C_{60} [in red, Inset (a)], as a function of the inverse number n of C_{60} layers. Here the reference is the gap of a fullerene in an infinite C_{60} monolayer [see Inset (b)]. The infinite size layer(s) limit is obtained by extrapolating disk(s) to their infinite R_{2D} radius. The dashed line represents the $[a/n + b]$ fit.

We finally conclude this study by making a connection to periodic slab calculations, namely the traditional approach for the study of surfaces with periodic boundary conditions. Along that line, we converge the polarization energy through the addition of C_{60} infinite sublayers, rather than by increasing the radius R of a half-sphere of polarizable molecules. Such a representation can be obtained by creating stacks of disks with increasing lateral radius R_{2D}, as represented in the Inset Fig. 8(a), extrapolating the polarization energy to infinity for a given number (n) of layers with an asymptotic scaling low in $1/R_{2D}^2$ (see Ref. 53). Taking as a reference an infinite size monolayer, as illustrated in the Inset Fig. 8(b), the polarization energy scales as $(1/n)$. We emphasize that for

(n=5), only 85% of the polarization energy is captured. Summing the 0.85 eV polarization energy of one C_{60} embedded in an infinite size monolayer, and the asymptotic polarization energy of 0.54 eV coming from an infinite number of layers with respect to a single monolayer, we find a total polarization energy of 1.39 eV. This value is nearly identical to the 1.40 eV value obtained with the "half-sphere" surface approach.

C. Discussion on CPU and memory requirements

The calculation for the biggest studied system made of 3367 C_{60}, amounting to 202 020 atoms as discussed in section III B, required around 8000 total CPU hours distributed on 720 cores. Alternatively, this represent a typical wall-time (time to completion of the run) of approximately 11 hours. Such a calculation was performed using the Irene SKL partition (Intel Skylake 8168 processors, with a base frequency of 2.70 GHz) of the IRENE supercomputer from GENCI-IDRIS. Only 2.5 terabytes of memory were used to study this system, making possible to run such calculation on smaller size computer clusters. Such a small memory footprint was made possible thanks to the small dimension (60) of the polarization basis for each fragment in the environment. Let's stress out here that describing each fullerene by its full def2-TZVP-RI basis, of dimension 5700, would lead to a memory footprint of 2.6 petabytes for each related matrices, with similar requirements for the total Coulomb potential $\mathbf{V}$ or the total susceptibility matrix $\mathbf{X}$. It would have been impossible to store any of these big matrices on the 310 terabytes available on the supercomputer used for this study. The present numbers are certainly indicative and may change depending on the chosen parameters, but they illustrate how efficient QM/QM (GW/RPA) calculations can be when using compressed susceptibility blocks for the environment.

IV. CONCLUSION

We have presented a fully *ab initio* QM/QM embedded *GW* calculation with a polarizable environment containing up to 200 000 atoms, with total typical CPU timings below 10 000 hours at the def2-TZVP level. Such calculations are made possible by adopting a fragment, or sub-system, approximation. Further, the susceptibility matrices associated with the fragments in the environment are reduced on-the-fly to a very low-rank representation, with block dimensions equivalent to the number of atoms in the fragment. This approach yields to representations even more compact than

standard semi-empirical approaches based on polarizable atoms described by onsite $\{p_x, p_y, p_z\}$ local induced dipoles, namely 1 degree of freedom per site as compared to the 3 degrees of polarizable atoms. Such a reduction allows inverting the Dyson equation for the screened Coulomb potential W with very limited CPU and memory requirements.

The present scheme is certainly far from being optimal. The choice of the same localization parameter (ζ) for all (s,p,d,f,g) diffuse channels in the ensemble of test functions can e.g. lead to forthcoming improvements. Our main point was however to show that very simple choices could already dramatically help in constructing model susceptibility operators both extremely compact and accurate in reproducing medium-to-long-range reaction fields.

As future directions, the scheme described here above can be merged with subsystem-DFT techniques[90–92] used to improve the input Kohn-Sham eigenstates, allowing to exploit fragment approximations while accounting for the frozen density of the neighboring molecules. In particular, the electrostatic field generated by the environment in the ground-state, of crucial importance in organic media composed of molecules with a permanent dipole, quadrupole, etc.[10,11,93–95] can be further accounted for at the DFT level.

Further, the block-diagonal form of the susceptibility can also be improved within a cluster expansion technique,[43] calculating the gas phase susceptibility of pairs of interacting neighboring fragments, building a susceptibility matrix that is tridiagonal by blocks rather than strictly diagonal. Within the scope of the present study, we tentatively assess the quality of the fragment approximation in the Supplementary Material, section III, where we compare, in the limit of a small number of fullerenes, the impact of neglecting the overlap of molecular orbitals between fragments. Allowing the delocalization of molecular orbitals over up to 4 fullerenes in the environment, we find that the fragment approximation does not induce errors larger than the percent on the calculated polarization energy. While we cannot study much larger systems at the many-body level without the fragment approximation, the 4-fullerenes environment induces a polarization correction to the central C_{60} gap that represents already ~30% of the true infinite surface limit.

Finally, the *GW* formalism allows to define a dynamical dielectric response, going beyond the standard low-frequency limit in the optical range of common PCM or QM/MM approaches. The importance of such dynamical corrections needs to be assessed, calculating the polarization energies within the full *GW* formalism rather than its static COHSEX limit, generalizing the recent merging of the *GW* formalism with a dynamical PCM.[24]

SUPPLEMENTARY MATERIAL

See the Supplementary Material for a study of (I) the error induced by the fit method on the HOMO and LUMO energies, (II) the impact of adding diffuse functions in the Kohn-Sham and associated auxiliary basis sets, and (III) a test of the fragment approach validity.

ACKNOWLEDGMENTS

DA is indebted to ENS Paris-Saclay for his PhD fellowship. This work was performed using HPC resources from GENCI-IDRIS (Grant 2021-A0110910016). XB and ID acknowledge support from the French Agence Nationale de la Recherche (ANR) under contract ANR-20-CE29-0005.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available within the article and its associated Supplementary Material.

Appendix A: Computation of the model susceptibility

In this Appendix, we compute the model susceptibility $\widetilde{\mathbf{X}}_g^{(\mathrm{I})}(\omega)$ solution of the equation 16. From now on and for compactness, we dropped the exponent (I) because we focus only on one fragment, and the frequency index ω. Results can be computed separately for each required frequency.

Keeping the same notations as the ones used in section II, we denote $\mathbf{B}$ the Coulomb matrix between the auxiliary basis and the test basis, such that its coefficients are $(Q\|t)$, where $(\cdot\|\cdot)$ indicates a Coulomb integral. We define $\mathbf{\Gamma}$ the Coulomb matrix between the polarization basis and the test basis, with coefficients $(\gamma\|t)$, and $\mathbf{E}$ (respectively $\mathbf{R}$) the overlap matrix between the auxiliary basis set (respectively the polarization basis) and the constraint basis, with coefficients $\langle Q|x^m y^n z^p\rangle$ (respectively $\langle\gamma|x^m y^n z^p\rangle$). The equation 16 can be rewritten

$$\widetilde{\mathbf{X}}_g = \underset{\widetilde{\mathbf{X}}}{\operatorname{argmin}} \left\| \mathbf{B}^\dagger \mathbf{X}_g \mathbf{B} - \mathbf{\Gamma}^\dagger \widetilde{\mathbf{X}} \mathbf{\Gamma} \right\|^2, \tag{A1}$$

with $\|\cdot\|$ the Frobenius norm, under constraints like the equation 17, which can also be rewritten

$$\mathbf{R}^\dagger \widetilde{\mathbf{X}} \mathbf{R} = \mathbf{E}^\dagger \mathbf{X}_g \mathbf{E}. \tag{A2}$$

To find this $N \times N$ matrix $\widetilde{\mathbf{X}}_g$, we build feasible solutions, namely matrices which satisfy all constraints enforced by the equation A2. Then, among all such feasible solutions, we compute the optimal one, which minimizes Eq. A1.

For compactness, we define $\mathbf{O} = \mathbf{E}^\dagger \mathbf{X}_g \mathbf{E}$ and $\mathbf{Z} = \mathbf{B}^\dagger \mathbf{X}_g \mathbf{B}$. Writing the compact singular value decomposition (SVD) of $\mathbf{R} = \mathbf{U}_\mathbf{R} \mathbf{\Sigma}_\mathbf{R} \mathbf{V}_\mathbf{R}^\dagger$, and assuming that the rank of $\mathbf{R}$ is equal to the number of constraints N_{cstr}, we search for solutions such that

$$\mathbf{U}_\mathbf{R}^\dagger \widetilde{\mathbf{X}} \mathbf{U}_\mathbf{R} = \mathbf{\Sigma}_\mathbf{R}^{-1} \mathbf{V}_\mathbf{R}^\dagger \mathbf{O} \mathbf{V}_\mathbf{R} \mathbf{\Sigma}_\mathbf{R}^{-1}.$$

Inverting $\mathbf{U}_\mathbf{R}$ requires to consider the nullspace of $\mathbf{U}_\mathbf{R}^\dagger$, of which we write an orthonormal basis as the columns of the $N \times (N - N_{\text{cstr}})$ matrix $\mathbf{K}$ such that $\mathbf{U}_\mathbf{R}^\dagger \mathbf{K} = \mathbf{0}$. Under these considerations, $\widetilde{\mathbf{X}}$ has the form

$$\widetilde{\mathbf{X}} = \mathbf{U}_\mathbf{R} \mathbf{\Sigma}_\mathbf{R}^{-1} \mathbf{V}_\mathbf{R}^\dagger \mathbf{O} \mathbf{V}_\mathbf{R} \mathbf{\Sigma}_\mathbf{R}^{-1} \mathbf{U}_\mathbf{R}^\dagger + \mathbf{K}\mathbf{M}_1 + \mathbf{M}_2 \mathbf{K}^\dagger.$$

So far, there are some redundancy in the $\mathbf{M}_1$ and $\mathbf{M}_2$ terms which can be lifted by projection on $\mathbf{U}$ and $\mathbf{K}$ supplementary subspaces. At the end, defining $\mathbf{C} = \mathbf{\Sigma}_\mathbf{R}^{-1} \mathbf{V}_\mathbf{R}^\dagger \mathbf{O} \mathbf{V}_\mathbf{R} \mathbf{\Sigma}_\mathbf{R}^{-1}$, and using the fact that both $\mathbf{O}$ and $\mathbf{Z}$ are symmetric, we find that

$$\widetilde{\mathbf{X}} = \mathbf{U}_\mathbf{R} \mathbf{C} \mathbf{U}_\mathbf{R}^\dagger + \mathbf{K}\mathbf{A}_1 \mathbf{K}^\dagger + \mathbf{U}_\mathbf{R} \mathbf{A}_2 \mathbf{K}^\dagger + \mathbf{K}\mathbf{A}_2^\dagger \mathbf{U}_\mathbf{R}^\dagger, \tag{A3}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are computed via the equation A1.

To solve this equation, let write $\mathbf{M}^\oplus$ the pseudo-inverse of the matrix $\mathbf{M}$. Let $\mathbf{P}$ be the matrix such that $\mathbf{P} = \left(\mathbf{K}^\dagger \mathbf{\Gamma}\right)^\oplus \mathbf{K}^\dagger$, let $\mathbf{I}_N$ be the identity matrix of size $N \times N$ and let $\mathbf{L}$ be the matrix such that $\mathbf{L} = \mathbf{R}^\oplus [\mathbf{I}_N - (\mathbf{\Gamma}\mathbf{P})]$. Using the formula of $\widetilde{\mathbf{X}}$ of Eq. A3 in Eq. A1, and keeping in mind that $\mathbf{U}$ and $\mathbf{K}$ span supplementary subspaces, the derivation condition with respect to $\mathbf{A}_1$ and $\mathbf{A}_2$ leads to

$$\boxed{\widetilde{\mathbf{X}}_g = \mathbf{L}^\dagger \mathbf{O} \mathbf{L} + \mathbf{P}^\dagger \mathbf{Z} \mathbf{\Gamma}^\oplus + (\mathbf{\Gamma}^\oplus)^\dagger \mathbf{Z} \mathbf{P} - \mathbf{P}^\dagger \mathbf{Z} \mathbf{P}}. \tag{A4}$$

If no constraints is enforced, we have $\mathbf{R} = \mathbf{L} = \mathbf{0}$, $\mathbf{K} = \mathbf{I}_N$ and $\mathbf{P} = \mathbf{\Gamma}^\oplus$. The formula simplifies thus into

$$\widetilde{\mathbf{X}}_g = (\mathbf{\Gamma}^\oplus)^\dagger \mathbf{Z} \mathbf{\Gamma}^\oplus. \tag{A5}$$

Appendix B: Definition of the test basis

We detail here how the test basis $\{t\}$ is set-up. It is in particular designed to reproduce as best as possible the effects of the susceptibility of reference at short to middle range. To do so, we rely on

test sets that sample uniformly the Coulomb interaction: starting from a set of functions $\{t_0\}$, we orthonormalize this set with respect to the Coulomb norm, leading to the test basis $\{t_1\}$. Namely, writing $\mathbf{V}_0$ the Coulomb matrix such that its coefficients are equal to $(t_0\|t'_0)$, where $(\cdot\|\cdot)$ indicates a Coulomb integral, the i^{th} vector's coordinates of $\{t_1\}$, in the basis set $\{t_0\}$, is given by the i^{th} column of $\mathbf{V}_0^{-1/2}$. This basis set is such that $(t_1\|t'_1) = \delta_{t_1,t'_1}$.

We create two of such sets: i) one spanning the auxiliary basis, namely using the notations of the section II $\{t_0\} = \{P\}$, and ii) another test basis $\{t_1^d\}$ based on the set of diffuse orbitals $\{t_0^d\}$. In this study, we used for $\{t_0^d\}$ a set of atom-centered (*s,p,d,f,g*) diffuse Gaussian orbitals with, for sake of simplicity, the same $e^{-\zeta r^2}$ radial part.

The final test basis $\{t\}$ is the direct sum of $\{t_1\}$ and $\{t_1^d\}$. The first set is down-weighted with respect to the second one, here by a factor $1/50$, to strengthen the influence of the diffuse functions during the fitting process.

Appendix C: Optimization of the polarization basis

In this Appendix, we explain how we compute the polarization basis $\{\gamma\}$ (see notation of section II), in which is computed $\widetilde{\mathbf{X}}_g$. More exactly, we develop our method to compute the $\{C_{\gamma P}\}$ coefficient set of Eq. 19. Starting from the equation A3, these coefficients were computed through two sub-problems.

1. Optimization for $N = N_{\text{cstr}}$

Keeping the same notation as in Appendix A, we started by studying the specific problem of $N = N_{\text{cstr}}$, with $\{\gamma_0\} = \{P\}$. To be more explicit, we re-index by 0 all matrices which depends directly on $\{\gamma_0\}$. $\mathbf{C}_0$ being invertible, with $\text{rank}(\mathbf{C}_0) = N_{\text{cstr}}$, there is a matrix $\mathbf{S}$ such that $\mathbf{A}_1 = \mathbf{S}\mathbf{C}_0\mathbf{S}^\dagger$, and $\mathbf{A}_2 = \mathbf{C}_0\mathbf{S}^\dagger$. We search for $\mathbf{S}$ such that

$$\mathbf{S}_{\text{opt}}^{(1)} = \underset{\mathbf{S}}{\text{argmin}} \left\| \mathbf{Z} - \mathbf{\Gamma}_0^\dagger (\mathbf{U}_{\mathbf{R}_0} + \mathbf{K}_0\mathbf{S})\mathbf{C}_0(\mathbf{U}_{\mathbf{R}_0} + \mathbf{K}_0\mathbf{S})^\dagger \mathbf{\Gamma}_0 \right\|.$$

The solution $\mathbf{S}_{\text{opt}}^{(1)}$ of this non-linear optimization problem is computed by a gradient descent algorithm.

We keep $\mathbf{U}_{\mathbf{R}} = \mathbf{U}_{\mathbf{R}_0} + \mathbf{K}_0\mathbf{S}_{\text{opt}}^{(1)}$ as a subset of the researched optimal directions $\{\gamma\}$. More exactly, the i^{th} column of $\mathbf{U}_{\mathbf{R}}$ corresponds to the coefficients $\{C_{\gamma P}\}$ of the i^{th} vector of the required basis set $\{\gamma\}$.

2. Optimization problem for $\mathbf{A}_2 = \mathbf{0}$

To find the other coefficients $\{C_{\gamma P}\}$, we use a greedy strategy. We consider only the first and second terms of the left-hand side of the equation A3, so namely $\mathbf{A}_2 = \mathbf{0}$, and write the factorized form of the general symmetric matrix $\mathbf{A}_1 = \mathbf{S}\boldsymbol{\Delta}\mathbf{S}^\dagger$, with $\boldsymbol{\Delta}$ a diagonal matrix. Independently of $\boldsymbol{\Delta}$, the required coefficients $\{C_{\gamma P}\}$ defining the optimal directions are only determined by the $\mathbf{S}$ matrix. Using the N_{cstr} first directions we have found, given $\mathbf{Z}_{\text{cstr}} = \mathbf{Z} - \boldsymbol{\Gamma}_0^\dagger \mathbf{U}_\mathbf{R} \mathbf{C}_0 \mathbf{U}_\mathbf{R}^\dagger \boldsymbol{\Gamma}_0$, and $m = \max(0, N - N_{\text{cstr}})$, we thus search for solutions $\mathbf{S}_{\text{opt}}^{(2)}$ of the low rank sub-problem

$$\mathbf{S}_{\text{opt}}^{(2)} = \underset{\mathbf{S},\boldsymbol{\Delta}\,/\,\operatorname{rank}(\mathbf{S})\leq m}{\operatorname{argmin}} \left\| \mathbf{Z}_{\text{cstr}} - \boldsymbol{\Gamma}_0^\dagger \mathbf{K}_0 \mathbf{S}\boldsymbol{\Delta}\mathbf{S}^\dagger \mathbf{K}_0^\dagger \boldsymbol{\Gamma}_0 \right\|.$$

Defining $\mathbf{J} = \mathbf{K}_0^\dagger \boldsymbol{\Gamma}_0$, and using its SVD $\mathbf{J} = \mathbf{U}_\mathbf{J} \boldsymbol{\Sigma}_\mathbf{J} \mathbf{V}_\mathbf{J}^\dagger$, this problem is equivalent to

$$\mathbf{S}_{\text{opt}}^{(2)} = \underset{\mathbf{S},\boldsymbol{\Delta}\,/\,\operatorname{rank}(\mathbf{S})\leq m}{\operatorname{argmin}} \left\| \mathbf{V}_\mathbf{J}^\dagger \mathbf{Z}_{\text{cstr}} \mathbf{V}_\mathbf{J} - \boldsymbol{\Sigma}_\mathbf{J} \mathbf{U}_\mathbf{J}^\dagger \mathbf{S}\boldsymbol{\Delta}\mathbf{S}^\dagger \mathbf{U}_\mathbf{J} \boldsymbol{\Sigma}_\mathbf{J} \right\|. \tag{C1}$$

Using the SVD of $\mathbf{V}_\mathbf{J}^\dagger \mathbf{Z}_{\text{cstr}} \mathbf{V}_\mathbf{J} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^\dagger$, and defining $\boldsymbol{\Sigma}_1^{(m)}$ the diagonal matrix made of the first m singular values of $\boldsymbol{\Sigma}_1$, the Eckart–Young–Mirsky theorem leads to a solution $\mathbf{S}_{\text{opt}}^{(2)}$ of Eq. C1 such that

$$\mathbf{S}_{\text{opt}}^{(2)} = \left[\sqrt{\boldsymbol{\Sigma}_1^{(m)}} \mathbf{V}_1^\dagger \boldsymbol{\Sigma}_\mathbf{J}^{-1} \mathbf{U}_\mathbf{J}^\dagger \right]^\dagger, \quad \boldsymbol{\Delta} = \mathbf{V}_1^\dagger \mathbf{U}_1.$$

The required coefficients $\{C_{\gamma P}\}$ are given by the columns of $\mathbf{K}_0 \mathbf{S}_{\text{opt}}^{(2)}$.

At the end of the process, the coordinates of the N polarization vectors $\{\gamma\}$ in the auxiliary basis, corresponding to the $\{C_{\gamma P}\}$ coefficients, are given by the columns of $\mathbf{U}_\mathbf{R}$ and of $\mathbf{K}_0 \mathbf{S}_{\text{opt}}^{(2)}$. In the specific case where no constraints is enforced, $\mathbf{R}_0$ is not defined, nor $\mathbf{U}_{\mathbf{R}_0}$, and $\mathbf{K}_0$ is the identity matrix. So the basis $\{\gamma\}$ is given by the columns of $\mathbf{S}_{\text{opt}}^{(2)}$, with $\mathbf{Z}_{\text{cstr}} = \mathbf{Z}$ and $\mathbf{J} = \boldsymbol{\Gamma}_0$.

We emphasize that such a method can be easily adapted to another choice of the first guess $\{\gamma_0\}$, different of the auxiliary basis.

REFERENCES

[1] M. Born, Z. Phys. **1**, 45 (1920).

[2] J. Jackson, Classical electrodynamics (Wiley; 2nd edition, 1975) Chap. Multipoles, Electrostatics of Macroscopic Media, Dielectrics.

[3] S. Miertuš, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).

[4]E. Cancès, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107**, 3032 (1997), https://doi.org/10.1063/1.474659.

[5]M. A. Thompson and G. K. Schenter, J. Phys. Chem. **99**, 6374 (1995), https://doi.org/10.1021/j100017a017.

[6]A. Osted, J. Kongsted, K. V. Mikkelsen, P.-O. Åstrand, and O. Christiansen, J. Chem. Phys. **124**, 124503 (2006), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2176615/16711271/124503_1_online.pdf.

[7]Lin and J. Gao, J Chem. Theory Comput. **3**, 1484 (2007), pMID: 26633219, https://doi.org/10.1021/ct700058c.

[8]C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, and B. Mennucci, J. Chem. Theory Comput. **5**, 1838 (2009), pMID: 26610008, https://doi.org/10.1021/ct9001366.

[9]D. Loco, L. Lagardère, O. Adjoua, and J.-P. Piquemal, Acc. Chem. Res. **54**, 2812 (2021), pMID: 33961401, https://doi.org/10.1021/acs.accounts.0c00662.

[10]G. D'Avino, L. Muccioli, C. Zannoni, D. Beljonne, and Z. G. Soos, J. Chem. Theory Comput. **10**, 4959 (2014), pMID: 26584380, https://doi.org/10.1021/ct500618w.

[11]G. D'Avino, L. Muccioli, F. Castet, C. Poelking, D. Andrienko, Z. G. Soos, J. Cornil, and D. Beljonne, J. Phys. Condens. Matter **28**, 433002 (2016).

[12]L. Hedin, Phys. Rev. **139**, A796 (1965).

[13]G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980).

[14]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[15]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[16]W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).

[17]G. Csanak, H. Taylor, and R. Yaris, Adv. At. Mol. Phys **7**, 287 (1971).

[18]G. Strinati, Riv. Nuovo Cimento **11**, 1 (1988).

[19]S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998).

[20]M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).

[21]L. X. Benedict, E. L. Shirley, and R. B. Bohn, Phys. Rev. Lett. **80**, 4514 (1998).

[22]I. Duchemin, D. Jacquemin, and X. Blase, J. Chem. Phys. **144**, 164106 (2016), https://doi.org/10.1063/1.4946778.

[23]I. Duchemin, C. A. Guido, D. Jacquemin, and X. Blase, Chem. Sci. **9**, 4430 (2018).

[24]J. M. Clary, M. Del Ben, R. Sundararaman, and D. Vigil-Fowler, J. Appl. Phys. **134**, 085001 (2023), https://pubs.aip.org/aip/jap/article-

pdf/doi/10.1063/5.0160173/18093460/085001_1_5.0160173.pdf.

[25]B. Baumeier, M. Rohlfing, and D. Andrienko, J. Chem. Theory Comput. **10**, 3104 (2014), pMID: 26588281, https://doi.org/10.1021/ct500479f.

[26]J. Li, G. D'Avino, I. Duchemin, D. Beljonne, and X. Blase, J. Phys. Chem. Lett. , 2814 (2016), pMID: 27388926, https://doi.org/10.1021/acs.jpclett.6b01302.

[27]J. Li, G. D'Avino, I. Duchemin, D. Beljonne, and X. Blase, Phys. Rev. B **97**, 035108 (2018).

[28]J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Çaylak, Y. Khalak, P. Madhikar, G. Tirimbò, and B. Baumeier, J. Chem. Theory Comput. **14**, 6253 (2018), pMID: 30404449, https://doi.org/10.1021/acs.jctc.8b00617.

[29]H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).

[30]D. Foerster, P. Koval, and D. Sánchez-Portal, J. Chem. Phys **135**, 074105 (2011).

[31]D. Neuhauser, Y. Gao, C. Arntsen, C. Karshenas, E. Rabani, and R. Baer, Phys. Rev. Lett. **113**, 076402 (2014).

[32]P. Liu, M. Kaltak, J. Klimeš, and G. Kresse, Phys. Rev. B **94**, 165109 (2016).

[33]V. Vlček, E. Rabani, D. Neuhauser, and R. Baer, J. Chem. Theory Comput. **13**, 4997 (2017), pMID: 28876912, https://doi.org/10.1021/acs.jctc.7b00770.

[34]V. Vlček, W. Li, R. Baer, E. Rabani, and D. Neuhauser, Phys. Rev. B **98**, 075107 (2018).

[35]J. Wilhelm, D. Golze, L. Talirz, J. Hutter, and C. A. Pignedoli, J. Phys. Chem. Lett. **9**, 306 (2018), pMID: 29280376, https://doi.org/10.1021/acs.jpclett.7b02740.

[36]A. Förster and L. Visscher, J. Chem. Theory Comput. **16**, 7381 (2020), pMID: 33174743, https://doi.org/10.1021/acs.jctc.0c00693.

[37]M. Kim, G. J. Martyna, and S. Ismail-Beigi, Phys. Rev. B **101**, 035139 (2020).

[38]A. Kutepov, Comput. Phys. Commun. **257**, 107502 (2020).

[39]I. Duchemin and X. Blase, J. Chem. Theory Comput. **17**, 2383 (2021), pMID: 33797245, https://doi.org/10.1021/acs.jctc.1c00101.

[40]J. Wilhelm, P. Seewald, and D. Golze, J. Chem. Theory Comput. **17**, 1662 (2021), pMID: 33621085, https://doi.org/10.1021/acs.jctc.0c01282.

[41]T. Fujita and Y. Noguchi, Phys. Rev. B **98**, 205140 (2018).

[42]T. Fujita, Y. Noguchi, and T. Hoshi, J. Chem. Phys. **151**, 114109 (2019), https://doi.org/10.1063/1.5113944.

[43]T. Fujita and Y. Noguchi, J. Phys. Chem. A **125**, 10580 (2021), pMID: 34871000, https://doi.org/10.1021/acs.jpca.1c07337.

[44]J. Tölle, T. Deilmann, M. Rohlfing, and J. Neugebauer, J. Chem. Theory Comput. **17**, 2186 (2021), pMID: 33683119, https://doi.org/10.1021/acs.jctc.0c01307.

[45]G. Weng and V. Vlček, J. Chem. Phys. **155**, 054104 (2021), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0058410/14111299/054104_1_online.pdf.

[46]G. Weng, A. Pang, and V. Vlček, J. Phys. Chem. Lett. **14**, 2473 (2023), pMID: 36867592, https://doi.org/10.1021/acs.jpclett.3c00208.

[47]J. B. Neaton, M. S. Hybertsen, and S. G. Louie, Phys. Rev. Lett. **97**, 216405 (2006).

[48]Z.-F. Liu, F. H. da Jornada, S. G. Louie, and J. B. Neaton, J. Chem. Theory Comput. **15**, 4218 (2019), pMID: 31194538, https://doi.org/10.1021/acs.jctc.9b00326.

[49]Z.-F. Liu, J. Chem. Phys. **152**, 054103 (2020), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5140972/15569206/054103_1_online.pdf.

[50]F. Xuan, Y. Chen, and S. Y. Quek, J. Chem. Theory Comput. **15**, 3824 (2019), pMID: 31084031, https://doi.org/10.1021/acs.jctc.9b00229.

[51]K. Andersen, S. Latini, and K. S. Thygesen, Nano Lett. **15**, 4616 (2015), pMID: 26047386, https://doi.org/10.1021/acs.nanolett.5b01251.

[52]K. T. Winther and K. S. Thygesen, 2D Mater. **4**, 025059 (2017).

[53]D. Amblard, G. D'Avino, I. Duchemin, and X. Blase, Phys. Rev. Materials **6**, 064008 (2022).

[54]F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).

[55]B. Farid, *Electron Correlation in the Solid State - Chapter 3*, edited by N. March (Imperial College Press, London, 1999).

[56]G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[57]R. Martin, L. Reining, and D. Ceperley, *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, 2016).

[58]Y. Ping, D. Rocca, and G. Galli, Chem. Soc. Rev. **42**, 2437 (2013).

[59]D. Golze, M. Dvorak, and P. Rinke, Front. Chem. **7**, 377 (2019).

[60]I. Duchemin and X. Blase, J. Chem. Theory Comput. **16**, 1742 (2020), pMID: 32023052, https://doi.org/10.1021/acs.jctc.9b01235.

[61]I. Duchemin and X. Blase, J. Chem. Theory Comput. **17**, 2383 (2021), pMID: 33797245, https://doi.org/10.1021/acs.jctc.1c00101.

[62]O. Vahtras, J. Almlöf, and M. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).

[63]X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. **14**, 053020 (2012).

[64]I. Duchemin, J. Li, and X. Blase, J. Chem. Theory Comput. **13**, 1199 (2017), pMID: 28094983, https://doi.org/10.1021/acs.jctc.6b01215.

[65]M. Pavanello, J. Chem. Phys. **138**, 204118 (2013), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4807059/13279854/204118_1_online.pdf.

[66]J. Tölle and J. Neugebauer, J. Phys. Chem. Lett. **13**, 1003 (2022), pMID: 35061387, https://doi.org/10.1021/acs.jpclett.1c04023.

[67]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[68]F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).

[69]H. F. Wilson, F. Gygi, and G. Galli, Phys. Rev. B **78**, 113303 (2008).

[70]M. Govoni and G. Galli, J. Chem. Theory Comput. **11**, 2680 (2015), pMID: 26575564, https://doi.org/10.1021/ct500958p.

[71]J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996), https://doi.org/10.1063/1.472933.

[72]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999), https://doi.org/10.1063/1.478522.

[73]F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, J. Chem. Phys. **152**, 224108 (2020), https://doi.org/10.1063/5.0004608.

[74]F. Neese, WIREs Comput. Mol. Sci. **12**, e1606 (2022), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1606.

[75]I. Duchemin, J. Li, and X. Blase, J. Chem. Theory Comput. **13**, 1199 (2017), pMID: 28094983, https://doi.org/10.1021/acs.jctc.6b01215.

[76]P. A. Heiney, J. E. Fischer, A. R. McGhie, W. J. Romanow, A. M. Denenstein, J. P. McCauley Jr., A. B. Smith, and D. E. Cox, Phys. Rev. Lett. **66**, 2911 (1991).

[77]In the fragment approximation, band dispersion originating from wavefunction hybridization between fragments cannot be accounted for.

[78]T. Rangel, S. M. Hamed, F. Bruneval, and J. B. Neaton, J. Chem. Theory Comput. **12**, 2834 (2016), pMID: 27123935, https://doi.org/10.1021/acs.jctc.6b00163.

[79]F. Kaplan, M. E. Harding, C. Seiler, F. Weigend, F. Evers, and M. J. van Setten, J. Chem. Theory Comput. **12**, 2528 (2016), pMID: 27168352, https://doi.org/10.1021/acs.jctc.5b01238.

[80]F. Bruneval and M. A. L. Marques, J. Chem. Theory Comput. **9**, 324 (2013), pMID: 26589035, https://doi.org/10.1021/ct300835h.

[81]F. Jensen, J. Chem. Phys. **117**, 9234 (2002), https://pubs.aip.org/aip/jcp/article-pdf/117/20/9234/10843810/9234_1_online.pdf.

[82]D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3484283/15431477/134105_1_online.pdf.

[83]E. L. Shirley and S. G. Louie, Phys. Rev. Lett. **71**, 133 (1993).

[84]B. Reihl, Mater Res Soc Symp Proc **359**, 377 (1994).

[85]J. Weaver, J. Phys. Chem. Solids **53**, 1433 (1992).

[86]P. J. Benning, D. M. Poirier, T. R. Ohno, Y. Chen, M. B. Jost, F. Stepniak, G. H. Kroll, J. H. Weaver, J. Fure, and R. E. Smalley, Phys. Rev. B **45**, 6899 (1992).

[87]R. W. Lof, M. A. van Veenendaal, B. Koopmans, H. T. Jonkman, and G. A. Sawatzky, Phys. Rev. Lett. **68**, 3924 (1992).

[88]T. Takahashi, S. Suzuki, T. Morikawa, H. Katayama-Yoshida, S. Hasegawa, H. Inokuchi, K. Seki, K. Kikuchi, S. Suzuki, K. Ikemoto, and Y. Achiba, Phys. Rev. Lett. **68**, 1232 (1992).

[89]See the NIST website: https://webbook.nist.gov/chemistry/.

[90]C. R. Jacob and J. Neugebauer, WIREs Comput. Mol. Sci. **4**, 325 (2014), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1175.

[91]L. E. Ratcliff, W. Dawson, G. Fisicaro, D. Caliste, S. Mohr, A. Degomme, B. Videau, V. Cristiglio, M. Stella, M. D'Alessandro, S. Goedecker, T. Nakajima, T. Deutsch, and L. Genovese, J. Chem. Phys. **152**, 194110 (2020), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0004792/16718083/194110_1_online.pdf.

[92]W. Dawson, S. Mohr, L. E. Ratcliff, T. Nakajima, and L. Genovese, J. Chem. Theory Comput. **16**, 2952 (2020), pMID: 32216343, https://doi.org/10.1021/acs.jctc.9b01152.

[93]D. Varsano, E. Coccia, O. Pulci, A. M. Conte, and L. Guidoni, Comput. Theor. Chem. **1040-1041**, 338 (2014).

[94]C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Würthner, K. Meerholz, K. Leo, and D. Andrienko, Nat. Mater **14**, 434 (2015).

[95]J. Li, I. Duchemin, O. M. Roscioni, P. Friederich, M. Anderson, E. Da Como, G. Kociok-Köhn, W. Wenzel, C. Zannoni, D. Beljonne, X. Blase, and G. D'Avino, Mater. Horiz. **6**, 107 (2019).

Supplementary Material : Many-body *GW* calculations with very large scale polarizable environments made affordable: a fully ab initio QM/QM approach

David Amblard,[1] Xavier Blase,[1] and Ivan Duchemin[2]

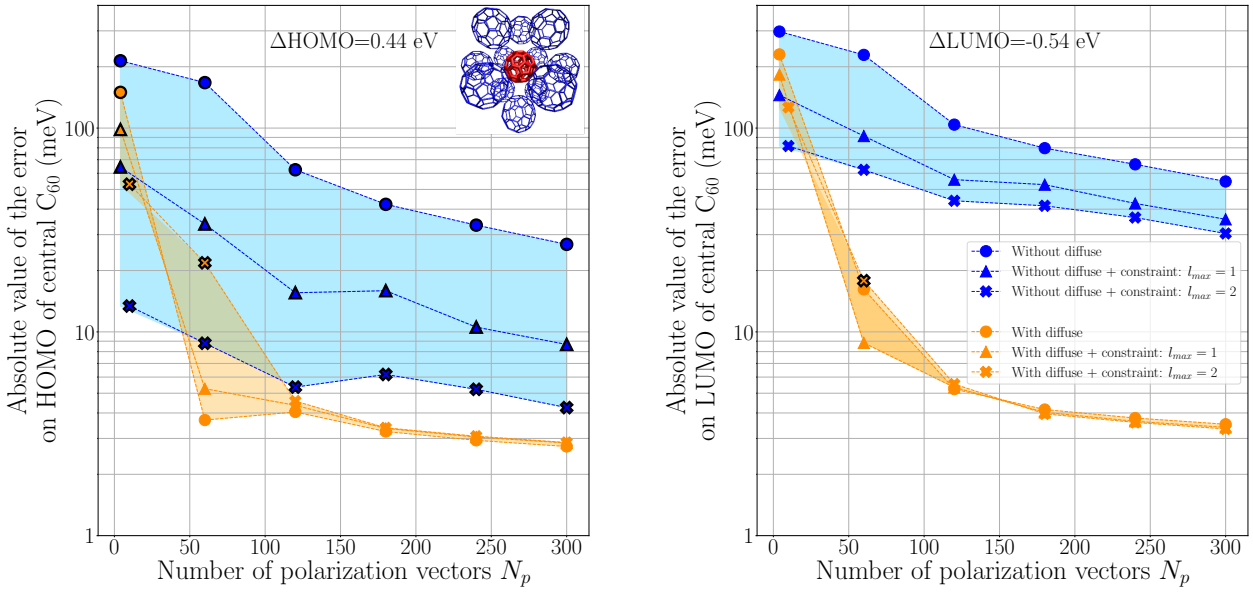
[1)] *Univ. Grenoble Alpes, CNRS, Inst NEEL, F-38042 Grenoble, France*

[2)] *Univ. Grenoble Alpes, CEA, IRIG-MEM-L_Sim, 38054 Grenoble, France*[a)]

[a)]Electronic mail: ivan.duchemin@cea.fr

I. ERROR INDUCED BY THE FIT METHOD FOR THE HOMO AND LUMO ABSOLUTE ENERGIES

In section III.A. of the main manuscript (Fig. 4), we study the error, as a function of the size N_p of the polarization basis per C_{60}, associated with the gap of one C_{60} embedded in a sphere of 12 neighbors. We focus now on the related HOMO and LUMO energies (Fig. S1). To allow a log-scale representation, only the absolute value of the errors are plotted. Negative errors are indicated by markers surrounded by black bold edges. Taking $N_p = 180$ polarization vectors, the value at which the error on the gap becomes lower than 1 meV, we find that the error on the absolute HOMO and LUMO energies is only slightly larger, at the 3-4 meV scale. As such, not only energy differences, but also absolute energy levels, are described quite accurately with very few optimal polarization vectors to describe the environment susceptibility.



(a) Error on HOMO

(b) Error on LUMO

FIG. S1: Same as Fig. 4 (main manuscript) but for the ΔLUMO and ΔHOMO energies, where the Δ means the evolution between the gas phase and the first-shell-of-neighbors environment [see Inset in (a)]. Data with black bold edges indicate negative values.

II. IMPACT OF AUGMENTING THE KOHN-SHAM AND AUXILIARY BASIS SETS

In the main manuscript, the def2-TZVP/def2-TZVP-RI Kohn-Sham and auxiliary basis sets have been used. We explore the impact of using basis sets containing more diffuse channels known to provide improved polarizability values for a given basis size.[S1,S2] To remain consistent with the def2-triple-zeta basis sets used in the main manuscript, we now adopt the def2-TZVPD/def2-TZVPD-RI basis sets that includes additional diffuse channels designed for the calculation of response properties.[S2] We reproduce in Fig. S2 the equivalent of the Fig. 4 of the main manuscript, showing very similar results. In the presence of diffuse orbitals ($\zeta = 0.01$ bohr^{-2}) in the test set, together with imposing the $l_{max} = 1,2$ constraints, the error reaches the meV error for N_p larger than $\sim$120. The converged reduction of the gap ΔGap $= -1.0$ eV is slightly larger than the ΔGap $= -0.98$ eV obtained with the def2-TZVP/def2-TZVP-RI basis sets, indicating a slight enhancement of the polarizability in the presence of diffuse channels.

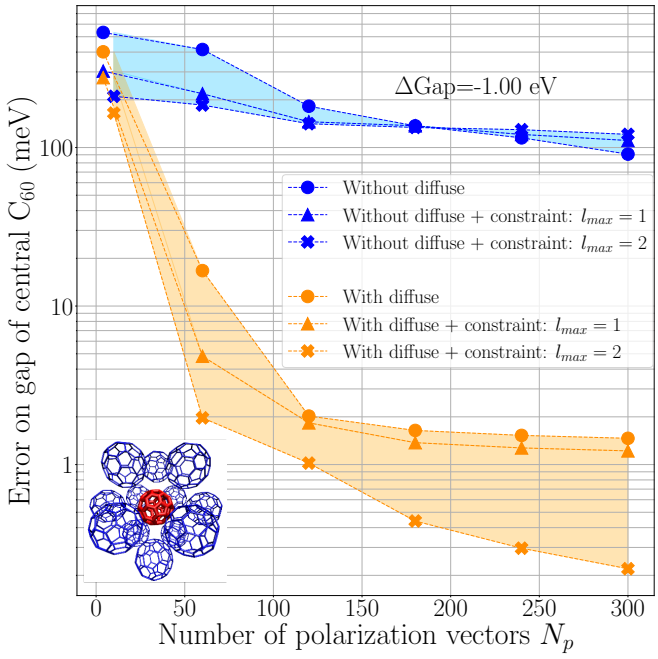
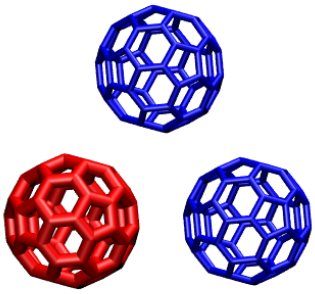


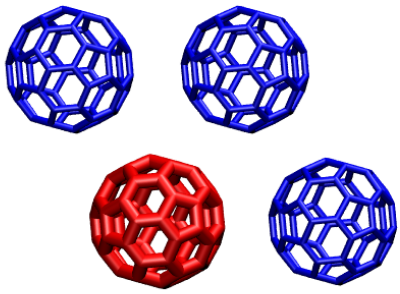
FIG. S2: Error on the central C_{60} [red, Inset] gap as a function of the number N_P of polarization vectors per C_{60} in the first-neighbors shell [blue, Inset]. Energies on the ordinates are in meV and log-scale. Results for test functions with and without diffuse orbitals, with and without constraints (see section II.C. of the main manuscript), are shown. The reference ΔGap $= -1.00$ eV corresponds to the gap reduction from the gas phase to the 13-C_{60} cluster when all fragments are described by their full auxiliary basis def2-TZVPD-RI ($N_p = 6900$).

III. VALIDITY OF THE FRAGMENT APPROACH

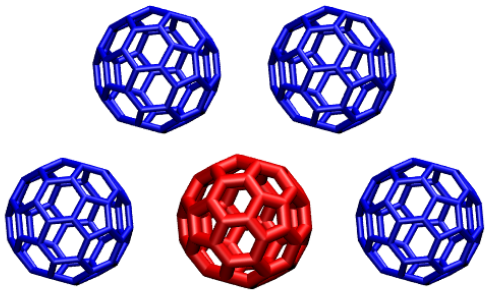
In this section, we compare calculations performed using the fragment approximation to reference calculations where all fullerenes in the environment are treated as a single subsystem, allowing wavefunction delocalization over all molecules in the environment. We again focus on the closing of the HOMO-LUMO gap of a "central" C_{60} (in red, Fig. S3) induced by the screening originating from the environment (blue molecules, Fig. S3). Due to the cost of calculating the susceptibility beyond the fragment approximation, the number of fullerenes in the environment is limited to four. To compare data on the same footing, we leave the central (red) fullerene as an isolated fragment to prevent delocalization of the associated HOMO and LUMO. What we want to explore is thus the impact of fragmenting the environment.



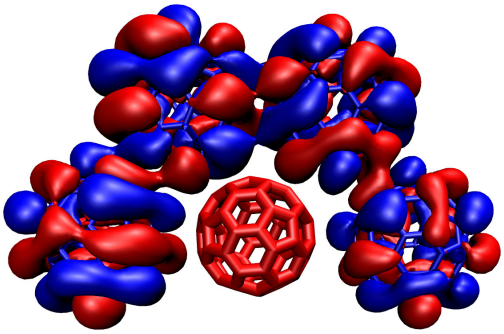
(a) 3 C_{60}



(b) 4 C_{60}



(c) 5 C_{60}



(d) HOMO $-$ 1

FIG. S3: Representations of the systems with (a) 3 C_{60}, (b) 4 C_{60} and (c) 5 C_{60} used to test the fragment approach. In red, the fullerene for which we calculate the difference ΔGap between the gap phase and the systems represented here. In (d) we represent the delocalized (HOMO $-$ 1) molecular orbital associated with the 4 C_{60} in the environment of system (c) beyond the fragment approximation.

The results are provided in Table S1. The impact of the fragment approximation, namely not allowing the delocalization of the molecular orbitals over all fullerenes in the environment, increases from 0.5 meV to 1.5 mev, remaining below the percent error. Even though limited to rather small systems, such an exploration provides a better understanding of the quality of the fragment approximation. Note that the ΔGap associated with the 4 C_{60} system (-0.45 eV) represents ∼32% of the infinite surface limit (ΔGap $= -1.4$ eV; see Fig. 7 main manuscript, blue symbols).

TABLE S1: Variation ΔGap, from the gas phase to the 3, 4, and 5 C_{60} systems, for the central C_{60} (in red in Fig. S3) allowing wavefunction delocalization over all C_{60} in the environment. The error induced by fragmenting the environment is provided in absolute value and in percentage. Calculations performed at the def2-TZVP/def2-TZVP-RI level (see details in the main text).

System	ΔGap (meV)	Error (meV)	Relative error (%)
3 C_{60}	-245	0.5	0.2
4 C_{60}	-350	2.2	0.6
5 C_{60}	-447	1.5	0.3

REFERENCES

[S1]F. Jensen, "Polarization consistent basis sets. III. The importance of diffuse functions," J. Chem. Phys. **117**, 9234–9240 (2002), https://pubs.aip.org/aip/jcp/article-pdf/117/20/9234/10843810/9234_1_online.pdf.

[S2]D. Rappoport and F. Furche, "Property-optimized Gaussian basis sets for molecular response calculations," J. Chem. Phys. **133**, 134105 (2010), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3484283/15431477/134105_1_online.pdf.