

Introducing DDEC6 atomic population analysis: part 5. New method to compute polarizabilities and dispersion coefficients

Thomas A. Manz^{1*}, Taoyi Chen¹, Daniel J. Cole², Nidia Gabaldon Limas¹, and Benjamin Fiszbein¹

¹Department of Chemical & Materials Engineering, New Mexico State University, Las Cruces, New Mexico, 88003-8001. ² School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK *Corresponding author email: tmanz@nmsu.edu

Abstract:

Polarizabilities and London dispersion forces are important to many chemical processes. Leading terms in these forces are often modeled using polarizabilities and C_n ($n=6, 8, 9, 10 \dots$) dispersion coefficients. Force fields for classical atomistic simulations can be constructed using atom-in-material dispersion coefficients and polarizabilities. This article addresses the key question of how to efficiently assign these parameters to constituent atoms in a material so that properties of the whole material are better reproduced. We develop a new set of scaling laws and computational algorithms (called MCLF) to do this in an accurate and computationally efficient manner across diverse material types. We introduce a conduction limit upper bound and m-scaling to describe the different behaviors of surface and buried atoms. We validate MCLF by comparing results to high-level benchmarks for isolated neutral and charged atoms, diverse diatomic molecules, various polyatomic molecules (e.g., polyacenes, fullerenes, and small organic and inorganic molecules), and dense solids (including metallic, covalent, and ionic). MCLF provides the non-directionally screened polarizabilities required to construct force fields, the directionally-screened static polarizability tensor components and eigenvalues, and environmentally screened C_6 coefficients. Overall, MCLF has improved accuracy and lower computational cost than the TS-SCS method. For TS-SCS, we compared charge partitioning methods and show DDEC6 partitioning yields more accurate results than Hirshfeld partitioning. MCLF also gives approximations for C_8 , C_9 , and C_{10} dispersion coefficients and Quantum Drude Oscillator parameters. For sufficiently large systems, our method's required computational time and memory scale linearly with increasing system size. This is a huge improvement over the cubic computational time of direct matrix inversion. As demonstrations, we study an ice crystal containing >250,000 atoms in the unit cell and the HIV reverse transcriptase enzyme complexed with an inhibitor molecule. This method should find widespread applications to parameterize classical force fields and DFT+dispersion methods.

1. Introduction

When combined with large-scale density functional theory calculations, the DDEC method has been shown to be suitable for assigning atom-centered point charges for flexible molecular mechanics force-field design.¹⁻⁶ The assignment of C_6 coefficients and atomic polarizabilities is another active area of research in force field design.^{3,7-10} Polarization effects are especially important for simulating materials containing ions.¹¹⁻¹⁸ When considered alongside the importance of accurate theoretical methods to study van der Waals interactions at the nanoscale,¹⁹ it is clear that a crucial feature of new methods to compute these important quantities is the ability to scale to large system sizes in reasonable computational time.

In this article, we develop new scaling laws and an associated method to compute polarizabilities and dispersion coefficients for atoms-in-materials (AIMs). These new scaling laws and computational method give good results for isolated atoms, diatomic molecules, polyatomic molecules, nanostructured materials, solids, and other materials. This new method is abbreviated MCLF according to the authors' last name initials (where the C is both for Chen and Cole). We performed tests on isolated neutral and charged atoms, small molecules, fullerenes, polyatomic molecules, solids, and a large biomolecule with MCLF. The results were compared with experimental data, high level CCSD calculations, time-dependent DFT (TD-DFT) calculations, or published force-field parameters.

As discussed in several recent reviews and perspectives, the dispersion interaction is a long-range, non-local interaction caused by fluctuating multipoles between atoms in materials.²⁰⁻²⁴ It is especially important in (i) condensed phases including liquids, supercritical fluids, solids, and colloids, (ii) nanostructure binding such as the graphene layers forming graphite, and (iii) the formation of noble gas dimers. The dispersion interaction is closely related to AIM polarizabilities. The dispersion energy can be described by an expansion series. The leading term is inversely proportional to R^6 , where R is the distance between two atoms. The coefficient of this term is called the C_6 dispersion coefficient, and this term quantifies fluctuating dipole-dipole interactions between two atoms. The intermolecular C_6 coefficient is given by the sum of all interatomic contributions

$$C_6^{\text{mol}} = \sum_{A \in M_1} \sum_{B \in M_2} C_{6AB}^{\text{eff}} \quad (1)$$

where M_1 and M_2 refer to the first and the second molecules. Higher-order terms represent different interactions.²⁵ For example, the eighth-order (C_8) term describes the fluctuating dipole-quadrupole interaction between two atoms. The ninth-order (C_9) term describes the fluctuating dipole-dipole-dipole interaction between three atoms. The tenth-order (C_{10}) term describes the fluctuating quadrupole-quadrupole and dipole-octupole interaction between two atoms.

Methods for computing polarizabilities and dispersion coefficients can be divided into two broad classes: (A) quantum chemistry methods that explicitly compute the system response to an electric field (e.g., time-dependent DFT (TD-DFT), CCSD perturbation response theory, etc.) and (B) AIM models. Class A methods can be highly accurate for computing polarizabilities and dispersion coefficients of a whole molecule, but they do not provide AIM properties. Therefore, class A methods cannot be regarded as more accurate versions of class B methods. Parameterizing a molecular mechanics force-field requires an AIM (i.e., class B) model. Our goal here is not merely to develop a computationally cheaper method than TD-DFT or CCSD perturbation response theory to compute accurate system polarizabilities and dispersion coefficients, but rather to exceed the capabilities of both of those methods by providing accurate AIM properties for force-field parameterization.

There are several existing frameworks for calculating AIM polarizabilities and/or dispersion coefficients. Applequist²⁶ introduced a formalism that represents the molecular polarizability tensor in terms of AIM polarizabilities via the dipole interaction tensor. Thole²⁷ refined this formalism by replacing atomic point dipoles with shape functions to avoid infinite interaction energies between adjacent atoms. Applequist’s and Thole’s methods use empirical atomic polarizability fitting to reproduce observed polarizabilities of small molecules.^{26, 27} Mayer et al. added charge-charge and dipole-charge interaction terms to calculate more accurate polarizabilities of conducting materials.^{28, 29} Grimme et al. presented the D3 geometry-based method to calculate C_6 , C_8 , and C_9 dispersion coefficients and dispersion energies, but this method does not yield polarizabilities.⁹ The exchange-dipole model (XDM) is an orbital-dependent approach that yields AIM dipole, quadrupole, and octupole polarizabilities and C_6 , C_8 , and C_{10} dispersion coefficients.³⁰⁻³⁴ Several density-dependent XDM variants have been formulated.^{35, 36} In 2009, Tkatchenko and Scheffler introduced the TS method for isotropic AIM polarizabilities and C_6 coefficients.¹⁰ Both the XDM and TS methods yielded isotropic AIM polarizabilities rather than molecular dipole tensors.^{30-33, 35, 36} In both the XDM and TS methods, the AIM unscreened polarizability is scaled as

$$\alpha^{\text{TS}} = \alpha^{\text{ref}} \frac{\langle r^3 \rangle^{\text{AIM}}}{\langle r^3 \rangle^{\text{ref}}} \quad (2)$$

where “ref” refers to the reference value for an isolated neutral atom of the same chemical element, “AIM” refers to the partitioned atom-in-material value, and $\langle r^3 \rangle$ refers to the r-cubed radial moment. Both the XDM and TS methods originally used Hirshfeld³⁷ partitioning to compute the $\langle r^3 \rangle^{\text{AIM}}$ values.^{10, 31}

In 2012, Tkatchenko et al. introduced self-consistent screening (TS-SCS) via the dipole interaction tensor to yield the molecular polarizability tensor and screened C_6 coefficients.³⁸ The TS-SCS dipole interaction tensor used a quantum harmonic oscillator (QHO) model similar to that used by Mayer but extended over imaginary frequencies and omitting charge-dipole and charge-charge terms.^{28, 29, 38} That same article also introduced a multibody dispersion (MBD³⁸) energy model based on a coupled fluctuating dipole model (CFDM³⁹). The TS-SCS screened static polarizability and characteristic frequency for each atom are fed into the CFDM model to obtain the MBD energy.³⁸ The TS-SCS approach has advantages of yielding a molecular polarizability tensor and AIM screened polarizabilities and AIM screened C_6 coefficients using only the system’s electron density distribution as input.³⁸

The TS-SCS method has several key limitations and flaws. Two key assumptions of the TS-SCS method are: (i) for a specific chemical element, the unscreened atomic polarizability is proportional to the atom’s $\langle r^3 \rangle$ moment, and (ii) for a specific chemical element, the unscreened C_6 coefficient is proportional to the atom’s polarizability squared.^{10, 38} However, in their work these hypotheses were not directly tested.^{10, 38} Later, Gould tested these two assumptions and found them inaccurate for describing isolated neutral atoms placed in a confinement potential.⁴⁰ Hirshfeld partitioning was used in the TS-SCS method^{10, 38} to compute the $\langle r^3 \rangle^{\text{AIM}}$. Because Hirshfeld partitioning uses isolated neutral atoms as references,³⁷ the Hirshfeld method typically severely underestimates NAC magnitudes.⁴¹⁻⁴⁴ The TS-SCS method assumes a constant unscreened polarizability-to- $\langle r^3 \rangle$ ratio and constant characteristic frequency (wp) for all charge states of a chemical element,³⁸ but these assumptions are not realistic. Due to these assumptions, the TS and TS-SCS methods are inaccurate for systems with charged atoms.⁴⁵⁻⁴⁸ Bucko et al. showed the TS and TS-SCS methods severely overestimate polarizabilities for dense solids.⁴⁸ The TS-SCS method has also not been optimized to work with conducting materials, and we show in Section 6

below the TS and TS-SCS methods sometimes predict erroneous polarizabilities even greater than for a perfect conductor. As discussed in Sections 4 and 6 below, the TS-SCS method also overestimates directional alignment of fluctuating dipoles at large interatomic distances. We also show the TS-SCS method sometimes gives asymmetric AIM polarizability tensors and unphysically negative AIM polarizabilities. Finally, the TS-SCS method is computationally expensive with required computational time scaling approximately cubically with increasing number of atoms in the unit cell.

Several research groups developed improvements to the TS-SCS method. Ambrosetti et al. introduced range separation to avoid double counting the long-range interactions in TS-SCS followed by MBD (aka MBD@rsSCS).⁴⁹ MBD@rsSCS improves the accuracy of describing directional alignments of fluctuating dipoles at large interatomic distances.⁴⁹ Bucko et al. used Iterative Hirshfeld (IH)⁴² partitioning in place of Hirshfeld partitioning to compute $\langle r^3 \rangle^{\text{AIM}}$.^{46, 48} While this was an improvement, it did not address several problems mentioned above. For example, the TS-SCS(IH) method still unrealistically assumed the unscreened polarizability-to- $\langle r^3 \rangle$ ratio is the same for various charge states of a chemical element.^{46, 48} This assumption was removed in the subsequent Fractionally Ionic (FI) method.⁴⁵ However, the FI approach requires separate reference polarizabilities and C_6 dispersion coefficients for all charge states of a chemical element.⁴⁵ This is extremely problematic, because some anions that exist in condensed materials (e.g., O^{2-}) have unbound electrons in isolation.⁵⁰ Although methods to compute charge-compensated reference ion densities have been developed,^{44, 46, 50-55} those methods do not presently extend to computing charge-compensated polarizabilities and C_6 coefficients of ions. Thus, several problems with the TS-SCS approach have not been satisfactorily resolved in the prior literature.

In Gould et al.'s FI method, reference free atom polarizabilities were computed for various whole numbers of electrons and interpolated to find fractionally charged free atom reference polarizabilities.⁴⁵ This yields different polarizability-to- $\langle r^3 \rangle$ ratios for different charge states of the same chemical element.⁴⁵ Due to the instability of highly charged anions, the -1 states of halogens were the only anions that Gould et al. computed self-consistently.⁴⁵ For other anions, Gould et al. resorted to using DFT orbitals from the neutral atoms to build non-self-consistent anions for polarizability and C_6 calculations,⁵⁶ but this severely underestimates the diffuseness of anions (i.e., underestimates their polarizabilities and C_6 coefficients). For example, O^- is more diffuse with a larger polarizability and C_6 than F^- , but Gould et al.⁵⁶ assigned the absurdly low $\alpha=5.40$ ($C_6=19.1$) to O^- compared to $\alpha=15.0$ ($C_6=73.5$) for F^- . This makes the FI method problematic for materials containing highly charged anions. Because FI was not included in VASP version 6b that we currently have access to, we did not perform FI calculations for comparison in this work.

In this article, we develop a new approach that resolves these issues. Our method uses DDEC6^{52, 57-59} partitioning to provide accurate NACs, atomic volumes, and radial moments as inputs. Our method has new scaling laws for the unscreened atomic polarizabilities, characteristic frequency (ω_p), and C_6 dispersion coefficients. The different scaling behaviors of surface and buried atoms are included via m-scaling. Our approach accurately handles the variability in polarizability-to- $\langle r^3 \rangle$ moment ratio for charged surface atoms while only requiring reference atomic polarizabilities, reference C_6 coefficients, and reference radial moments for isolated neutral atoms. It uses a new self-consistent screening procedure to compute screened polarizability tensors and C_6 coefficients. Our approach separates non-directional screening from directional screening of the dipole interaction tensor. This allows a conduction limit upper bound to be applied between non-directional and directional screening to ensure buried atoms do not have

a screened polarizability above the conduction limit. Our method yields three different types of dipole polarizabilities: (a) induced static polarizabilities corresponding to a uniform applied external electric field, (b) isotropic screened polarizabilities suitable as input into polarizable force-fields, and (c) fluctuating polarizabilities that are used to compute C_6 dispersion coefficients via the Casimir-Polder integral. When computing C_6 dispersion coefficients, we use multi-body screening to taper off the dipole directional alignment at large interatomic distances. To achieve a linearly scaling computational cost with increasing system size, the self-consistent screening is applied incrementally using an algorithm that avoids large matrix inversion. Richardson extrapolation provides high numeric precision. Through Quantum Drude Oscillator (QDO) parameterization, our method also yields higher-order polarizabilities (e.g., quadrupolar, octupolar, etc.) and higher-order dispersion coefficients (e.g., C_8 , C_9 , C_{10} , etc.) for AIMs. Other important improvements include: improved damping radii, proportional partitioning of off-diagonal polarizability components to avoid negative AIM polarizabilities, iterative update of the spherical Gaussian dipole width, and AIM polarizabilities are symmetric tensors.

Our method was designed to satisfy the following criteria:

- 1) The method should require only the system’s electron and spin density distributions as input;
- 2) The method should work for materials with 0, 1, 2, or 3 periodic boundary conditions;
- 3) The method should give accurate results for charged atoms in materials while only requiring reference polarizabilities and reference C_6 coefficients for neutral free atoms;
- 4) The method should give accurate results for diverse materials types: isolated atoms; small and large molecules; nanostructured materials; ionic, covalent, and metallic solids; etc.;
- 5) The method should give accurate results for both surface and buried atoms;
- 6) The method should yield both static polarizability tensors and polarizabilities suitable for constructing molecular mechanics force-fields;
- 7) The method should accurately describe both short- and long-range ordering of dipole polarizabilities and C_6 coefficients;
- 8) The method should have less than approximately 10% average error on C_6 coefficients and dipole polarizabilities for the benchmark sets studied here;
- 9) The method should include estimates for higher-order AIM polarizabilities (e.g., quadrupolar, octupolar, etc.) and dispersion coefficients (e.g., C_8 , C_9 , C_{10} , etc.);
- 10) The method should have low computational cost for both small and large systems.

There are two main applications for this MCLF method. First, the polarizabilities and dispersion coefficients can be used to partially parameterize molecular mechanics force-fields. In addition to polarizabilities and C_6 dispersion coefficients, those force-fields would also need to include net atomic charges (NACs), flexibility parameters (e.g., bond, angle, and torsion terms), exchange-repulsion parameters, (optionally) charge penetration parameters, and optionally other parameters. Second, the dispersion coefficients can be used to partially parameterize DFT+dispersion methods.²¹ In addition to the C_6 dispersion coefficients, an accurate DFT+dispersion method should also include higher-order dispersion (e.g., C_8 , C_9 , and/or C_{10} terms) or multi-body dispersion (MBD) combined with an accurate damping function.^{20, 21} (Partially analogous to the rsSCS@MBD method,⁴⁹ range separation would be required to avoid double counting dispersion interactions when combining a MCLF variant with a MBD

Hamiltonian.) Because density functional theory (DFT) and molecular mechanics are widely used in computational chemistry, our new method can have widespread applications.

The remainder of this article is organized as follows. Section 2 contains the background information. Section 3 contains the new isolated atom scaling law developed for MCLF. Section 4 describes the theory of the MCLF method. Section 5 describes the computational algorithm and timing results. Section 6 contains calculation results of C_6 coefficients and polarizabilities and comparisons to benchmark data. Section 7 is the conclusions.

2. Background Information

2.1 Benchmarking Methods

Experimental data and high-level quantum chemistry calculations were used as references in this work. Section 3.1 below describes reference polarizabilities and dispersion coefficients (C_6 , C_8 , and C_{10}) for the isolated atoms. For diatomic molecules in Section 6.1, we computed static polarizability tensors using CCSD calculations combined with the “polar” keyword in Gaussian09⁶⁰. As explained in Section 6.2 below, we set the reference static polarizability for dense solids to the lesser of the Clausius-Mossotti relation value and the conduction limit upper bound based on the experimental crystal structure geometry and dielectric constant.

For the small molecules in Section 6.3, reference static polarizabilities were obtained from published experiments. Experimental isotropic polarizabilities were extracted from dielectric constant or refractive index measurements having approximately 0.5% or less error.⁶¹ Refractive index can be measured by passing a light ray through a gas-phase sample.⁶² The polarizability $\alpha(\nu)$ of the sample at frequency ν can then be calculated using

$$\alpha(\nu) = \frac{3\Lambda}{4\pi} \frac{\eta^2(\nu) - 1}{\eta^2(\nu) + 2} - \frac{\mu^2}{3k_B T} \quad (3)$$

where η is the refractive index, μ is the dipole moment magnitude, T is absolute temperature, Λ is the volume per molecule, and k_B is Boltzmann’s constant.⁶¹ Static or low-frequency dielectric constants κ were obtained by measuring the ratio of the capacitance of a set of electrodes with the sample material in-between to the capacitance of the same electrodes with a vacuum in-between.⁶³ The polarizability of a gas-phase sample can then be calculated using the Clausius-Mossotti relation:⁶⁴

$$\alpha = \frac{3\Lambda}{4\pi} \frac{\kappa - 1}{\kappa + 2} \quad (4)$$

Reference C_6 coefficients for the atom/molecule pairs (Section 6.4) were taken from the experimentally extracted dipole oscillator strength distribution (DOSD) data of Meath and co-workers⁶⁵,⁶⁶ as tabulated in the supporting information of Bucko et al.⁴⁶

Time-dependent DFT (TD-DFT) and time-dependent Hartree-Fock (TD-HF) can be used to compute benchmark polarizabilities and dispersion coefficients. The Casimir-Polder integral is used to calculate C_6 coefficients from polarizabilities at imaginary frequencies (imfreqs).⁶⁷ For polyacenes (Section 6.5), reference C_6 coefficients and isotropic static polarizabilities are from the TD-DFT calculations of Marques et al.⁶⁸ For selected polyacenes, static polarizability tensor components were available as reference from Jiemchoorj et al.’s TD-DFT calculations.⁶⁹ Jiemchoorj et al. found their TD-DFT results were similar to TD-HF, experimental (where available), and CCSD (where available) results. For fullerenes (Section 6.5), the reference C_6 coefficients and isotropic static polarizabilities are

from the TD-HF calculations of Marques et al.⁶⁸ Marques et al. also obtained similar results using TD-DFT.⁶⁸

2.2 Notation

A system may have either 0, 1, 2, or 3 periodic boundary conditions. In periodic materials, the term ‘image’ refers to a translated image of the reference unit cell. Each image is designated by translation integers (L_1, L_2, L_3) that quantify the unit cell translation along the lattice vectors. The reference unit cell is the image designated by $(L_1, L_2, L_3) = (0, 0, 0)$. $-\infty \leq L \leq \infty$ along a periodic direction. $L = 0$ along a non-periodic direction. Similar to the notation previously used in the bond order article,⁵⁸ a capital letter (A, B, ...) designates an atom in the reference unit cell and a lowercase letter (a, b, ...) designates an image atom. For example, $b = (B, L_1, L_2, L_3)$ denotes a translated image of atom B.

Let \vec{R}_B represent the nuclear position of atom B in the reference unit cell. Then, the nuclear position of a translated image is

$$\vec{R}_b = \vec{R}_B + L_1 \vec{v}^{(1)} + L_2 \vec{v}^{(2)} + L_3 \vec{v}^{(3)} \quad (5)$$

where $\vec{v}^{(1)}$, $\vec{v}^{(2)}$, and $\vec{v}^{(3)}$ are the lattice vectors. The distance between the nuclear position of atom A and the translated image of atom B is

$$d_{Ab} = r^{AB,L} = \|\vec{R}_b - \vec{R}_A\| \quad (6)$$

Cartesian components ($\beta = x, y, z$) of the vector from atom A’s nuclear position to image b’s nuclear position are represented by

$$r_{\beta}^{AB,L} = \vec{R}_b - \vec{R}_A \quad (7)$$

A stockholder partitioning method assigns a set of atomic electron densities $\{\rho_A(\vec{r}_A)\}$ in proportion to atomic weighting factors $\{w_A(r_A)\}$

$$\rho_A(\vec{r}_A) = \rho(\vec{r}) w_A(r_A) / W(\vec{r}) \quad (8)$$

so that all sum to the total electron density

$$\rho(\vec{r}) = \sum_{A,L} \rho_A(\vec{r}_A) \quad (9)$$

$$W(\vec{r}) = \sum_{A,L} w_A(r_A) \quad (10)$$

where summation over A, L means summation over all atoms in the material. The number of electrons N_A and net atomic charge (q_A) assigned to atom A are

$$N_A = \oint \rho_A(\vec{r}_A) d^3 \vec{r}_A = \Theta_A - q_A \quad (11)$$

where Θ_A is the atomic number for atom A. As discussed in Section 2.4 below, different ways of defining $\{w_A(r_A)\}$ produce different stockholder methods.

\vec{r}_A is the vector from the image of atom A’s nuclear position to the spatial position \vec{r} :

$$\vec{r}_A = \vec{r} - L_1 \vec{v}^{(1)} - L_2 \vec{v}^{(2)} - L_3 \vec{v}^{(3)} - \vec{R}_A \quad (12)$$

The length of this vector is represented by

$$r_A = \|\vec{r}_A\| \quad (13)$$

The AIM radial moment of order ϕ is

$$\langle (r_A)^\phi \rangle = \oint (r_A)^\phi \rho_A(\vec{r}_A) d^3\vec{r}_A \quad (14)$$

$\langle r^2 \rangle$, $\langle r^3 \rangle$, and $\langle r^4 \rangle$ are short-hand for $\langle (r_A)^2 \rangle$, $\langle (r_A)^3 \rangle$, and $\langle (r_A)^4 \rangle$, respectively.

Post-stockholder methods require stockholder AIM properties as inputs (e.g., $\{\rho_A(\vec{r}_A)\}$, $\{\langle (r_A)^\phi \rangle\}$) and return properties such as bond orders, polarizabilities, dispersion coefficients, etc. A full method specification indicates both the post-stockholder method and the stockholder method. For example, TS(H) represents Tkatchenko-Scheffler unscreened scaling using Hirshfeld partitioning inputs.¹⁰ As a second example, Manz(DDEC6) and Manz(DDEC3) denote Manz bond orders computed using the DDEC6 and DDEC3 partitioning inputs, respectively.⁵⁸ This notation recognizes that: (a) the same post-stockholder method may be combined with different stockholder partitioning methods (e.g., TS-SCS(H), TS-SCS(IH), TS-SCS(DDEC6)) and (b) the same stockholder partitioning method may be combined with different post-stockholder methods (e.g., TS-SCS(DDEC6) and MCLF(DDEC6)). Because all MCLF results in this article used DDEC6 partitioning, herein we use the less precise but shorter term ‘MCLF’ in place of the full ‘MCLF(DDEC6)’ designation.

Calculating dispersion coefficients involves integrating polarizabilities over imfregs. This is inconvenient in two respects. First, it is easier to deal with real-valued variables rather than imaginary-valued ones. Second, numeric integration from zero to infinite imaginary frequency is inconvenient, because infinity cannot be readily divided into finite intervals for numeric integration. Letting ω represent an imaginary frequency magnitude, we used the following variable transformation to solve these two problems:

$$u = \frac{\text{Nimfregs}}{1 + \omega}, \quad \omega(u) = \frac{\text{Nimfregs}}{u} - 1 \quad (15)$$

This conveniently transforms integration limits $\omega = [0, \infty)$ into $u = [\text{Nimfregs}, 0)$, which upon changing the integrand’s sign gives integration limits $u = (0, \text{Nimfregs}]$. As shown in Section 5.2 below, this allows convenient Rhomberg integration using integration points $u = 1, 2, \dots, \text{Nimfregs}$. In this article, $\alpha(u)$ denotes the polarizability at the *imaginary frequency* whose magnitude equals $\omega(u)$.

2.3 Details of the TS-SCS methodology

Figure 1 is a flow diagram of the TS-SCS method. The supporting information of Bucko et al.⁴⁷ gave the step-by-step calculation of the self-consistent screening process. Here, we follow the presentation of Bucko et al., except using the variable substitution of eqn (15). For atoms A and B, they define a many-body polarizability matrix P, with its inverse Q having the form

$$Q_{\alpha\beta}^{AB} = \frac{1}{\alpha_A^{\text{unscreened}}} \delta_{AB} \delta_{\alpha\beta} + \tau_{\alpha\beta}^{AB} \quad (16)$$

where α, β designate Cartesian components. Square matrices P and Q have x, y, and z spatial indices for every atom to give a total of 3Natoms rows. The last term on the right-hand side is the dipole interaction tensor which has the form

$$\tau_{\alpha\beta}^{AB} = -\sum_L \frac{3r_{\alpha}^{AB,L}r_{\beta}^{AB,L} - (r^{AB,L})^2\delta_{\alpha\beta}}{(r^{AB,L})^5} \left(\operatorname{erf}\left(\frac{r^{AB,L}}{\sigma_{AB}(u)}\right) - \frac{2}{\sqrt{\pi}} \frac{r^{AB,L}}{\sigma_{AB}(u)} e^{-\left(\frac{r^{AB,L}}{\sigma_{AB}(u)}\right)^2} \right) \quad (17)$$

$$\sum_L \frac{4}{\sqrt{\pi}} \frac{1}{(\sigma_{AB}(u))^3} \frac{r_{\alpha}^{AB,L}r_{\beta}^{AB,L}}{(r^{AB,L})^2} e^{-\left(\frac{r^{AB,L}}{\sigma_{AB}(u)}\right)^2}$$

where summation over L means summation over all periodic translation images (if any). $\sigma_{AB}(u)$ is the attenuation length for the pair of atoms A and B

$$\sigma_{AB}(u) = \sqrt{\sigma_A^2(u) + \sigma_B^2(u)} \quad (18)$$

The spherical Gaussian dipole width is obtained from

$$\sigma_A(u) = \left(\sqrt{\frac{2}{\pi}} \frac{\alpha_A(u)}{3} \right)^{1/3} \quad (19)$$

and the isotropic dynamical atomic polarizability is

$$\alpha_A^{\text{unscreened}}(u) = \frac{\alpha^{\text{unscreened}}}{1 + (\omega(u)/wp)^2} \quad (20)$$

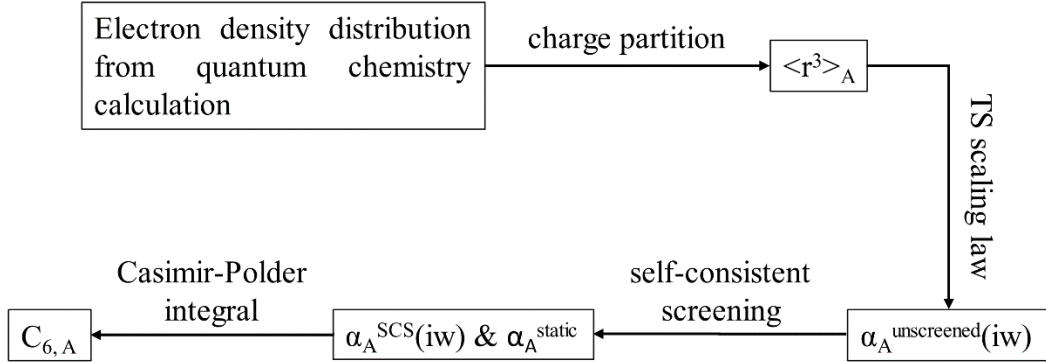


Figure 1: Flow diagram for TS-SCS method

In the TS and TS-SCS methods,

$$\alpha^{\text{unscreened}} = \alpha^{\text{TS}} \quad (21)$$

where α^{TS} is calculated by eq. (2). AIM polarizability tensors are computed using the partial contraction

$$\vec{\vec{\alpha}}_A^{\text{TS-SCS}}(u) = \sum_{B=1}^{\text{Natoms}} \mathbf{P}_{\alpha\alpha}^{AB}(u) \quad (22)$$

with the static polarizability tensor corresponding to $u = \text{Nimfreqs}$. The screened frequency-dependent isotropic polarizability is computed as one third of the trace of the three-by-three polarizability tensor obtained by partial contraction of P

$$\alpha_A^{\text{TS-SCS}}(\mathbf{u}) = \text{trace}(\vec{\vec{\alpha}}_A(\mathbf{u}))/3 = \frac{1}{3} \sum_{\alpha} \sum_{B=1}^N P_{\alpha\alpha}^{\text{AB}}(\mathbf{u}) \quad (23)$$

These are fed into the Casimir-Polder integral (eqn (74)) to compute C_6^{AA} .

2.4 Electron Density Partitioning Methods

In Hirshfeld partitioning introduced in 1977, atoms are partitioned to resemble the neutral atom references.³⁷ This makes the atoms tend to have lower charge than they should have.⁴¹ The iterative Hirshfeld partition (IH) keeps updating the reference with the charge state of the atom.⁴² However, this approach leads to the runaway charge problem in some cases.⁵² As shown in reference⁴⁸ and Section 6 below, using TS or TS-SCS with Hirshfeld or IH partitioning overestimates polarizabilities for dense solids.

Manz and Sholl presented DDEC1 and 2 atomic population analysis methods in 2010.⁵¹ By simultaneously optimizing the atomic density distributions to be close to spherical average density and to resemble charge-compensated reference ion densities, this method can give chemically meaningful NACs and accurate electrostatic potential for some materials, but was later found to give runaway charges in other materials. In 2012, Manz and Sholl presented the DDEC3 method that partially fixes the runaway charge problem by increasing the optimization landscape curvature via using conditioned reference densities and imposing an exponential decay constraint on each atom's electron density tail.⁴⁴

Manz and Limas presented DDEC6 partitioning in 2016.^{52, 57} In this method, the runaway charge problem is fixed by stop updating the reference ion charge after certain iterative steps. Also, new constraints are added to the decay rate of the buried atom tails. The weighted spherical average improves the effect of spherical averaging during charge partitioning. Along with the guaranteed convergence in seven steps, this method is very accurate, cost efficient and chemically meaningful.^{52, 57, 59} In 2017, Manz published a new method for computing bond orders, which is based on DDEC6.⁵⁸

3. New isolated atom scaling laws

3.1 Reference Data

The reference polarizabilities (α_{CCSD}) used in this work are our calculated polarizabilities using the CCSD method with def2QZVPPDD basis set. (The def2QZVPPDD basis set is defined in the ESI†). We tested two different methods: (a) using Gaussian09⁶⁰ keyword 'polar' to compute the molecular static polarizability tensor using perturbation response theory and (b) using Gaussian09 keyword 'field' to manually apply a small constant external electric field \vec{E} in order to calculate the molecular static polarizability tensor as $\vec{\vec{\alpha}} = \delta\vec{\mu}/\delta\vec{E}$ where $\vec{\mu}$ is the molecular dipole moment. However, many of the manual (i.e., keyword = 'field') calculations did not converge and the converged results were not as consistent with Gould and Bucko's data⁷⁰ as the perturbation response calculations. So we decided to use the perturbation response calculations (i.e., keyword = 'polar') for all elements except Y. For Y the keyword = 'field' polarizability was used, because the perturbation response calculation gave an unreasonably low polarizability of 88.98 compared to $\alpha_{\text{Gould}} = 163$ while the manually applied field polarizability of 158.81 was close to Gould's value and followed the trend of neighboring elements: $\alpha_{\text{Sr, CCSD}} = 204.51$ $\alpha_{\text{Zr, CCSD}} = 143.47$.

Figure 2 shows that our calculated polarizabilities are mostly consistent with Gould's. We used α_{CCSD} rather than α_{Gould} as the reference free atom polarizability, because our radial moments come from

the same CCSD calculations as used to compute α_{CCSD} . For elements using a relativistic effective core potential (RECP) in the def2-QZVPPDD basis set, the radial moments of core electrons replaced by the RECP are added back in using a reference core density library; thus yielding effective all-electron radial moments. Since Gould and Bucko used the aufbau principle for electron configurations of transition metal atoms, their calculations do not necessarily correspond to the ground state spin multiplicity for transition metal atoms.⁷⁰

CCSD in Gaussian09 does not have the capability of calculating C_6 coefficients. Therefore, to maximize consistency between the free atom reference radial moments, polarizabilities, and C_6 coefficients, our reference C_6 coefficients were calculated as

$$C_6^{\text{ref}} = C_6^{\text{Gould}} \left(\frac{\alpha_{\text{CCSD}}}{\alpha_{\text{Gould}}} \right)^{1.5} \quad (24)$$

Where α_{Gould} and C_6^{Gould} are Gould and Bucko's values using TD-DFT.⁵⁶ This C_6 rescaling makes C_6^{ref} corresponding to α_{CCSD} , which corresponds to the computed radial moments. The 3/2 power occurs in eqn (24), because α and C_6 for a free atom are approximately proportional to the free atom's effective radius to the fourth and sixth powers, respectively (see Table S1 of ESI†).

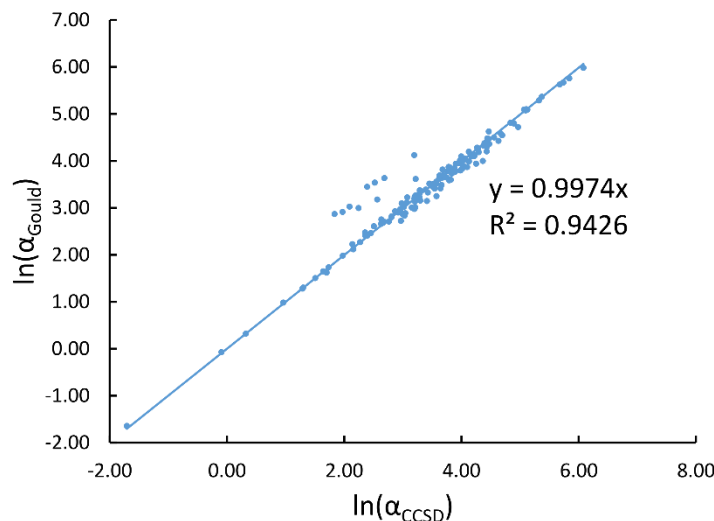


Figure 2: $\ln(\alpha_{\text{CCSD}})$ vs $\ln(\alpha_{\text{Gould}})$

The reference α , C_6 , wp, r moments, and damping radii are listed in the ESI†. This dataset contains neutral elements 1–86 except the f-block elements (58–71). The reason for excluding the f-block and heavier elements is the def2QZVPPDD basis set we are using does not include these elements. The dataset also includes +1 cations of elements 3–7, 11–17, 19–57 and 72–86 and -1 anions of F, Cl, Br, I, and At. These ions were also self-consistently calculated by Gould and Bucko⁵⁶ as well. Gould and Bucko included additional anions which were not computed self-consistently, and we omit these because self-consistent polarizabilities are unavailable.⁷⁰ The reference wp were calculated from the CCSD polarizability and the C_6^{ref} using⁷¹

$$wp = \frac{4}{3} \frac{C_6}{\alpha^2} \quad (25)$$

The reference C_8 and C_{10} are from Porsev et al.⁷² and Tao et al.⁷³ and are listed in the ESI†. This dataset contains H, Li, Na, K, Rb, Cs, He, Ne, Ar, Kr, Xe, Be, Mg and Ca.

3.2 Deriving the New Scaling Laws

Johnson and Becke assumed that for a given chemical element, the polarizability of atoms in a material should be proportional to the $\langle r^3 \rangle$ moment of the atom-in-material.³⁴ This assumption was subsequently adopted by Tkatchenko and Scheffler when formulating the TS method.¹⁰ Of course, this is not the same as assuming polarizabilities of isolated atoms across different chemical elements should be proportional to their $\langle r^3 \rangle$ moments. As pointed out by Gould, the isolated atom polarizabilities are not proportional to their $\langle r^3 \rangle$ moments.⁴⁰ Figure 3 is a plot of $\ln(\alpha_{\text{CCSD}})$ versus $\ln(\langle r^3 \rangle)$ for the isolated atoms. The plot shows a poor correlation with a $R^2 = 0.7706$ value indicating that isolated atom polarizability is only weakly correlated to $\langle r^3 \rangle$ moment. This motivated us to develop a new polarizability scaling law that applies both to isolated atoms and to atoms-in-materials across different chemical elements.

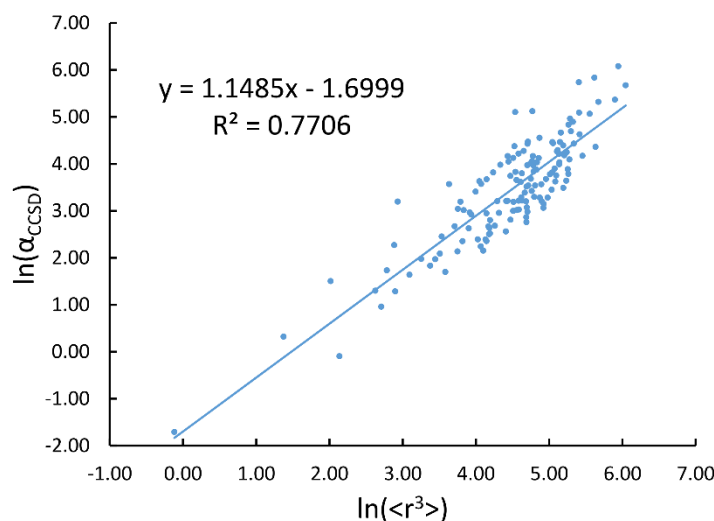


Figure 3: $\ln(\alpha_{\text{CCSD}})$ versus $\ln(\langle r^3 \rangle)$

We tested seven models containing electron number and different combinations of the r moments. Table 1 lists the R^2 values of the 7 models. The coefficients are obtained by least squares fitting of a linear combination of the log values of r moments and electron numbers to α , C_6 , or wp using a Matlab program we wrote. For example, cell (2, 2) in Table 1 is the R^2 value of 0.6626 obtained by fitting $\log(\langle r \rangle)$ and \log of electron number to $\log(\alpha_{\text{CCSD}})$. The results show that using only one r moment does not yield high R^2 value. Combinations of two or more r moments give higher R^2 values, with the $\langle r^3 \rangle$ & $\langle r^4 \rangle$ model giving the best average performance.

Table 2 lists parameters for the $\langle r^3 \rangle$ & $\langle r^4 \rangle$ model. The proposed relations between α , wp , and C_6 and the parameters have the following form:

$$\alpha = e^{-2.2833} N^{0.2892} \frac{\langle r^4 \rangle^{3.3372}}{\langle r^3 \rangle^{3.1657}} \quad (26)$$

$$wp = \frac{e^{1.6336} \langle r^3 \rangle^{3.7003}}{N^{0.3167} \langle r^4 \rangle^{3.2228}} \quad (27)$$

$$C_6 = e^{-3.2206} N^{0.2618} \frac{\langle r^4 \rangle^{3.4516}}{\langle r^3 \rangle^{2.6311}} \quad (28)$$

Figure 4, Figure 5, and Figure 6 show strong correlation between the model predicted α , C_6 , and wp and the reference data with R^2 values of 0.9549, 0.977, and 0.8494, respectively.

Table 1: R^2 values for fitted parameters using CCSD r moments

model	α_{CCSD}	C_6	wp
N & $\langle r \rangle$	0.6626	0.7619	0.3922
N & $\langle r^2 \rangle$	0.8021	0.8809	0.5489
N & $\langle r^3 \rangle$	0.8831	0.9414	0.6615
N & $\langle r^4 \rangle$	0.9242	0.9672	0.7317
N, $\langle r^2 \rangle$ & $\langle r^3 \rangle$	0.9457	0.973	0.8222
N, $\langle r^2 \rangle$ & $\langle r^4 \rangle$	0.9545	0.9772	0.8452
N, $\langle r^3 \rangle$ & $\langle r^4 \rangle$	0.9549	0.977	0.8494

Table 2: Parameter coefficients for the new scaling law

α		C_6		wp	
component	coefficient	component	coefficient	component	coefficient
constant	-2.2833	constant	-3.2206	constant	1.6336
N	0.2892	N	0.2618	N	-0.3167
$\langle r^3 \rangle$	-3.1657	$\langle r^3 \rangle$	-2.6311	$\langle r^3 \rangle$	3.7003
$\langle r^4 \rangle$	3.3372	$\langle r^4 \rangle$	3.4516	$\langle r^4 \rangle$	-3.2228

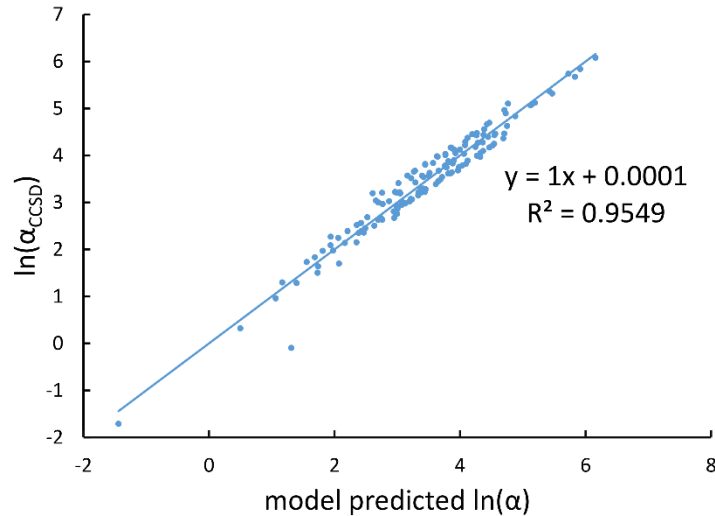


Figure 4: Model predicted $\ln(\alpha)$ vs reference $\ln(\alpha)$

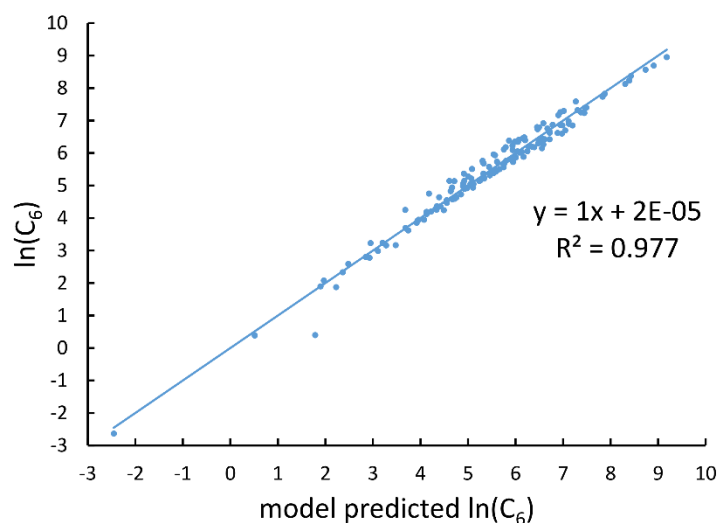


Figure 5: Model predicted $\ln(C_6)$ vs reference $\ln(C_6)$

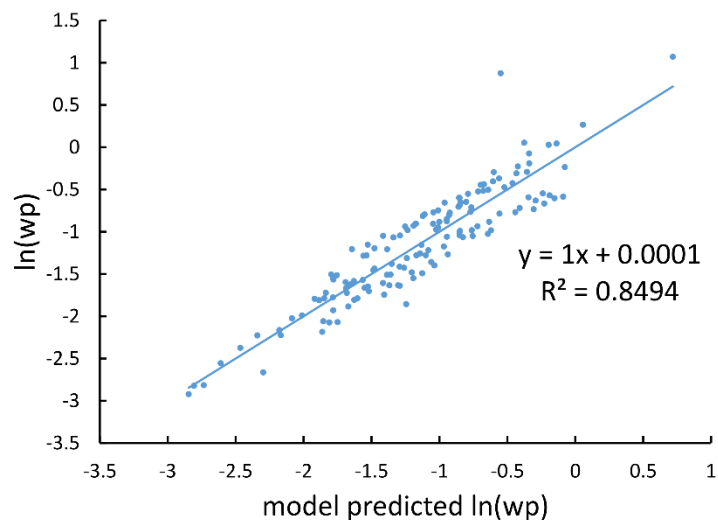


Figure 6: Model predicted $\ln(wp)$ vs reference $\ln(wp)$

To test the robustness and transferability of the different models, the following tests were performed as shown in Table 3. The “PW91 refitted” column are the R^2 values obtained by refitting the model parameters with r moments from PW91. The “PW91 predicted” column are the R^2 values calculated using CCSD model parameters from Table 1 but with PW91 r moments instead of CCSD r moments.

Table 3 shows that the $\langle r^3 \rangle$ & $\langle r^4 \rangle$ model has the highest R^2 in both tests; therefore, this model is the most robust and transferable.

Table 3: R^2 values for parameters using PW91 r moments

model	PW91 refitted			PW91 predicted		
	α_{CCSD}	C_6	wp	α_{CCSD}	C_6	wp
N, $\langle r^2 \rangle$ & $\langle r^3 \rangle$	0.8785	0.9130	0.7571	0.8127	0.8658	0.6181
N, $\langle r^2 \rangle$ & $\langle r^4 \rangle$	0.8904	0.9176	0.7942	0.8276	0.8706	0.6780
N, $\langle r^3 \rangle$ & $\langle r^4 \rangle$	0.8942	0.9181	0.8159	0.8345	0.8726	0.7104

Hence, the new polarizability scaling law for an isolated atom is

$$\alpha^{\text{unscreened}} = \alpha^{\text{ref}} \left(\frac{\langle r^3 \rangle^{\text{ref}}}{\langle r^3 \rangle^{\text{AIM}}} \right)^{3.1657} \left(\frac{\langle r^4 \rangle^{\text{AIM}}}{\langle r^4 \rangle^{\text{ref}}} \right)^{3.3372} \left(\frac{N^{\text{AIM}}}{N^{\text{ref}}} \right)^{0.2892} \quad (29)$$

where N is the number of electrons, the superscript “ref” means the value of neutral atom reference, and “AIM” means the value for atom-in-material after partitioning. The new wp scaling law for an isolated atom is

$$\text{wp}^{\text{unscreened}} = \text{wp}^{\text{ref}} \left(\frac{\langle r^3 \rangle^{\text{AIM}}}{\langle r^3 \rangle^{\text{ref}}} \right)^{3.7003} \left(\frac{\langle r^4 \rangle^{\text{ref}}}{\langle r^4 \rangle^{\text{AIM}}} \right)^{3.2228} \left(\frac{N^{\text{ref}}}{N^{\text{AIM}}} \right)^{0.3167} \quad (30)$$

$C_6^{\text{unscreened}}$ for an isolated surface atom is then computed via eqn (25). These scaling laws allow different charge states of an atom to be accurately described while using only reference polarizability and wp values for a neutral free atom of the same element. For α , the effective power of the effective radius is $4*3.3372 - 3*3.1657 = 3.8517$, which is approximately 4. For wp, the effective power of the effective radius is $3*3.7003 - 4*3.2228 = -1.7903$, which is approximately 4. Scaling laws for non-isolated atoms will be addressed in Section 4 below.

3.3 Higher Order Dispersion Coefficients and Quantum Drude Oscillator Parameters

In this section, we consider higher-order dispersion coefficients C_8 , C_9 , and C_{10} and their mixing rules. The contribution of the three-body C_9 term to the dispersion energy is typically less than 10%.⁹ Nevertheless, McDaniel et al.⁷⁴ showed that in order to obtain accurate results from force-field simulations for condensed phases, the three-body term (E^{ABC}) should be included. Tang et al. showed that the attractive potential at well depth for two free atoms is mainly composed of C_6 , C_8 , and C_{10} with contributions of roughly 65%, 25%, and 7% respectively.⁷⁵ The rest comes from higher-order terms. Because the C_8 , C_9 , and C_{10} terms have modest contributions, we decided to include them in our model.

The $C_{8,A}$ coefficient describes the fluctuating-dipole-fluctuating-quadrupole two-body dispersion interaction between atoms of the same type. We defined two dimensionless groups for least-squares fitting to obtain a model for $C_{8,A}$:

$$\text{group_1} = \ln \left[\left(C_{6,A}^{\text{non-dir}} \right)^{1/3} \left(\frac{\langle r^3 \rangle_A}{\langle r^4 \rangle_A} \right)^2 \right] \quad (31)$$

$$\text{group_2} = \ln \left[\frac{C_{8,A}}{C_{6,A}^{\text{non-dir}}} \left(\frac{\langle r^3 \rangle_A}{\langle r^4 \rangle_A} \right)^2 \right] \quad (32)$$

The reason for using $\langle r^3 \rangle$ and $\langle r^4 \rangle$ is that these are the same r moments used in models discussed above. Since $C_{8,A}$ describes the fluctuating dipole-quadrupole coupling while $C_{6,A}$ describes the fluctuating dipole-dipole coupling, there is no reason to believe directional effects on $C_{8,A}$ follow those on $C_{6,A}$. Therefore, our correlations for higher-order dispersion coefficients (i.e., C_8 , C_9 , and C_{10}) do not include directional coupling. $C_{6,A}^{\text{non-dir}}$ is obtained using the imfreq-dependent non-directionally screened atomic

polarizability $\alpha_A^{\text{non-dir}}(u)$ in the Casimir-Polder integral. Linear fitting was performed to obtain the slope and intercept for group 2 as a function of group 1. The results were 0.8305 for the slope and 1.7327 for the intercept yielding:

$$C_{8,A} = e^{1.7327} (C_{6,A}^{\text{non-dir}})^{1+(0.8305/3)} \left(\frac{\langle r^4 \rangle}{\langle r^3 \rangle} \right)^{2-2*0.8305} \quad (33)$$

The top panel of Figure 7 shows strong correlation between the model predicted $C_{8,A}$ and the reference data^{72, 73} for selected isolated atoms with MARE of 14.6%.

The Quantum Drude Oscillator (QDO) model provides a natural framework for describing multibody polarizability and dispersion interactions beyond the dipole approximation, including quadrupolar, octupolar, and high-order interactions.^{7, 76, 77} A QDO consists of a negative pseudoparticle coupled via a harmonic potential to a pseudonucleus.^{7, 76, 77} This harmonic coupling produces a Gaussian charge distribution.^{7, 76} In our model, one QDO is centered on each atom in the material. Each QDO is completely described by three parameters: (a) an effective mass (m_A^{QDO}), (b) an effective charge (q_A^{QDO}), and (c) an effective frequency (wp_A^{QDO}).^{7, 76} Using literature relations⁷ applied to our non-directionally screened quantities yields

$$C_{10,A} = \left(\frac{49}{40} \right) \frac{(C_{8,A})^2}{C_{6,A}^{\text{non-dir}}} \quad (34)$$

$$wp_A^{\text{QDO}} = \frac{4C_{6,A}^{\text{non-dir}}}{3\alpha_A^{\text{force-field}}} \quad (35)$$

$$m_A^{\text{QDO}} = \frac{15\alpha_A^{\text{force-field}}}{4C_{8,A}} \quad (36)$$

$$q_A^{\text{QDO}} = -\sqrt{\frac{20(C_{6,A}^{\text{non-dir}})^2}{3C_{8,A}}} \quad (37)$$

$$C_{9,ABC} = \frac{\alpha_A^{\text{force-field}} \alpha_B^{\text{force-field}} \alpha_C^{\text{force-field}} wp_A^{\text{QDO}} wp_B^{\text{QDO}} wp_C^{\text{QDO}} (wp_A^{\text{QDO}} + wp_B^{\text{QDO}} + wp_C^{\text{QDO}})}{2(wp_A^{\text{QDO}} + wp_B^{\text{QDO}})(wp_A^{\text{QDO}} + wp_C^{\text{QDO}})(wp_B^{\text{QDO}} + wp_C^{\text{QDO}})} \quad (38)$$

$$\alpha_A^{\text{quadrupole}} = \frac{3\alpha_A^{\text{force-field}} C_{8,A}}{20C_{6,A}^{\text{non-dir}}} \quad (39)$$

$$\alpha_A^{\text{octapole}} = \frac{\alpha_A^{\text{force-field}}}{20} \left(\frac{C_{8,A}}{C_{6,A}^{\text{non-dir}}} \right)^2 \quad (40)$$

$$C_{8,AB} = \frac{15}{2} wp_A^{\text{QDO}} wp_B^{\text{QDO}} \left(\frac{\alpha_A^{\text{force-field}} \alpha_B^{\text{quadrupole}}}{(wp_A^{\text{QDO}} + 2wp_B^{\text{QDO}})} + \frac{\alpha_B^{\text{force-field}} \alpha_A^{\text{quadrupole}}}{(2wp_A^{\text{QDO}} + wp_B^{\text{QDO}})} \right) \quad (41)$$

$$C_{10,AB} = 7wp_A^{\text{QDO}} wp_B^{\text{QDO}} \left[3 \frac{\alpha_A^{\text{force-field}} \alpha_B^{\text{octapole}}}{(wp_A^{\text{QDO}} + 3wp_B^{\text{QDO}})} + 3 \frac{\alpha_B^{\text{force-field}} \alpha_A^{\text{octapole}}}{(3wp_A^{\text{QDO}} + wp_B^{\text{QDO}})} + 5 \frac{\alpha_A^{\text{quadrupole}} \alpha_B^{\text{quadrupole}}}{(wp_A^{\text{QDO}} + wp_B^{\text{QDO}})} \right] \quad (42)$$

The $C_{9,ABC}$ QDO mixing rule⁷ is similar to that described much earlier by Tang using a Padé approximation.⁷¹ The bottom panel of Figure 7 shows strong correlation between the model predicted $C_{10,A}$ and the reference data^{72, 73} for selected isolated atoms with MARE of 18.6%.

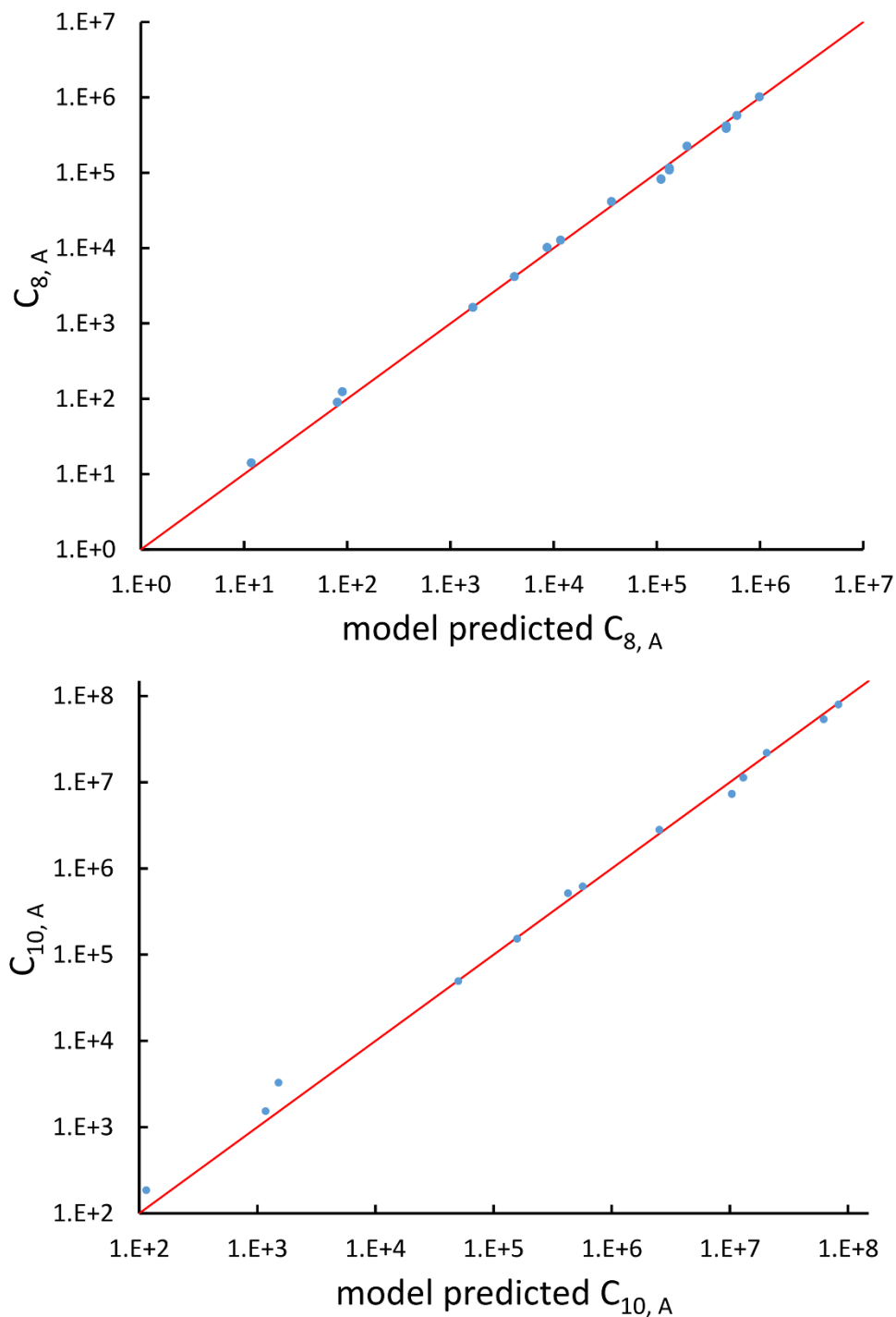


Figure 7: Model predicted C_{8A} and C_{10A} vs reference

When constructing a force-field using the MCLF dispersion coefficients, care should be taken not to double-count the three-body dipole-dipole-dipole interactions. Specifically, the MCLF directional

screening (i.e., C_6^{screened}) already includes the three-body dipole-dipole-dipole interactions for the calculated system. (These were included via the directional dipole interaction tensor which was used in turn to compute C_6^{screened} .) For example, if C_6^{screened} is computed using the MCLF method for a single benzene molecule, then the intramolecular dipole-dipole-dipole interactions are already included via the C_6^{screened} coefficient term, but the intermolecular dipole-dipole-dipole interactions are not already included and should be added in the force-field using the C_9 coefficient term. In this case, the force-field's C_9 three-body term should be constructed to include atom triplets from two or three molecules, but not from a single molecule.

4. Theory of MCLF Method

4.1 New Polarizability Component Partition

During tests, we noticed TS-SCS sometimes gives negative atomic polarizabilities. For example, in the ZrO molecule, the polarizability of oxygen is -0.607 from TS-SCS. This causes subsequent methods depending on the TS-SCS method to fail: (a) the C_6 coefficient from the Casimir polder integral will be unphysical, (b) corresponding vdW radii in the TS-SCS method will be complex (since they depend on the cubed root of the polarizability)¹⁰, and (c) MBD, rsSCS@MBD, and related methods require non-negative atomic polarizabilities as input^{38, 45, 49}. In TS-SCS, the partial contraction of P follows the assumed form of Applequist et al.²⁶:

$$P_{\alpha\beta}^A = \sum_B P_{\alpha\beta}^{AB} \quad (43)$$

This sometimes yields negative atomic polarizabilities, because the mixed contribution $P_{\alpha\beta}^{AB}$ (which might be negative) between an atom A with small polarizability and an atom B with large polarizability can surpass the magnitude of $P_{\alpha\beta}^{AA}$. In our new method, atoms with larger pre-screened polarizability get a proportionally larger piece of the screening mixed polarizability contribution:

$$\alpha_A^{\text{non-dir}}(u) = \sum_B P^{AB}(u) \left(\frac{2\alpha_A^{\text{unscreened}}(u)}{\alpha_A^{\text{unscreened}}(u) + \alpha_B^{\text{unscreened}}(u)} \right) \quad (44)$$

$$P_{\alpha\beta}^A(u) = \sum_B \left(\frac{\alpha_A^{\text{non-dir}}(u)}{\alpha_A^{\text{non-dir}}(u) + \alpha_B^{\text{non-dir}}(u)} (P_{\alpha\beta}^{AB}(u) + P_{\beta\alpha}^{AB}(u)) \right) \quad (45)$$

$$P_{\alpha\beta}^A = \sum_B \left(\frac{\alpha_A^{\text{force-field}}}{\alpha_A^{\text{force-field}} + \alpha_B^{\text{force-field}}} (P_{\alpha\beta}^{AB} + P_{\beta\alpha}^{AB}) \right) \quad (46)$$

Eqs. (44), (45), and (46) correspond to the non-directional, fluctuating, and static polarizabilities, respectively. After this new partition is applied, the oxygen in ZrO has the MCLF polarizability of 6.754. Also, Eq. (45) ensures the atom-in-material polarizability $P_{\alpha\beta}^A$ is a symmetric tensor just like the total polarizability tensors⁷⁸ of all molecules. In contrast, the TS-SCS polarizability tensor for an atom-in-material is sometimes asymmetric with respect to the spatial coordinates (e.g., the xy and yx components are different). As an example, the ESI[†] contains the TS-SCS and MCLF results files for dibromomethane, where the TS-SCS yz and zy polarizability components for the last Br atom were 4.05 and 7.31, respectively; the MCLF method gave 7.17 for both components.

The symmetric atomic polarizability tensor can be visualized by plotting the ellipsoid⁷⁸

$$\left(\frac{\vec{r}_A \cdot \hat{e}^I}{\lambda_1}\right)^2 + \left(\frac{\vec{r}_A \cdot \hat{e}^{II}}{\lambda_2}\right)^2 + \left(\frac{\vec{r}_A \cdot \hat{e}^{III}}{\lambda_3}\right)^2 = 1 \quad (47)$$

where λ_1 , λ_2 , and λ_3 are its three eigenvalues and \hat{e}^I , \hat{e}^{II} , and \hat{e}^{III} are the corresponding mutually orthogonal eigenvectors. Figure 8 plots atomic polarizability tensors for the carbonyl sulfide molecule. All three atoms showed enhanced polarizability along the bond direction. The atom-in-material polarizability along a unit direction \hat{k} is quantified by the projection $\hat{k} \cdot \vec{\alpha}_A \cdot \hat{k}$. Choosing \hat{k} parallel to the inter-nuclear direction gives the bond polarizability of two bonded atoms: $\hat{k} \cdot (\vec{\alpha}_A + \vec{\alpha}_B) \cdot \hat{k}$.⁷⁸ Of interest, Raman spectrum peak intensities are proportional to the change in projected polarizability as vibration occurs.⁷⁹⁻⁸¹

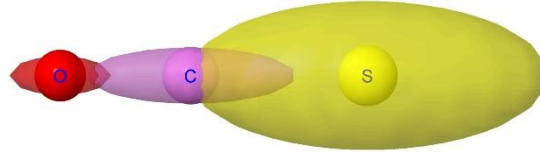


Figure 8: Illustration of MCLF atom-in-material polarizability tensors for the carbonyl sulfide molecule. The sulfur atom had much higher polarizability than the carbon and oxygen atoms. All three atoms showed enhanced polarizability along the bond direction. Only the relative sizes and shapes of the ellipsoids were drawn to scale.

4.2 Polarizability upper bound

Consider a perfectly conducting plate with thickness d in an external electric field E applied perpendicular to its surface. From Gauss' Law, the two faces perpendicular to E develop surface charge densities $\sigma' = \pm 2\epsilon_0 E$ and form a dipole moment $\mu = \sigma' d \cdot \text{Area} = 2\epsilon_0 E d \cdot \text{Area}$. Its polarizability is $p = \mu / E = 2\epsilon_0 d \cdot \text{Area}$, which in atomic units ($4\pi\epsilon_0 = 1$ in atomic units) gives $\alpha = p / (4\pi\epsilon_0) = d \cdot \text{Area} / (2\pi)$, also referred to as the polarizability volume since it carries volume units. Thus, the polarizability-to-volume ratio of the perfectly conducting plate is $1/(2\pi)$ in atomic units.

As a second example, consider a sphere of radius R and dielectric constant κ placed in a constant externally applied electric field E along the z -direction. This sphere will develop a dipole moment of⁸²

$$\mu^{\text{sphere}} = (4\pi\epsilon_0) R^3 \left(\frac{\kappa - 1}{\kappa + 2} \right) E \quad (48)$$

Therefore, its polarizability in atomic units is $R^3 (\kappa - 1) / (\kappa + 2)$, and its polarizability-to-volume ratio is $(3/(4\pi))((\kappa - 1)/(\kappa + 2))$ in atomic units. The solution for a perfect conductor can be obtained by taking the limit $\kappa \rightarrow \infty$ which yields $3/(4\pi)$ as the polarizability-to-volume ratio for the conducting sphere.

In theory, the polarizability caused by distortion of the electron cloud for fixed nuclear positions should be less than or equal to that of a perfect conductor. Comparing the above results for the conducting plate and conducting sphere shows the polarizability-to-volume ratio of a perfect conductor is shape-dependent; this is due to directional interactions within the material. To address this issue, we apply a conduction limit upper bound during non-directional screening, before any directional screening occurs.

Since the surface charge density of the conducting plate is uniform while that of the conducting sphere is not, and since stacking plates is space-filling while stacking spheres is not space-filling, the polarizability-to-volume ratio of the conducting plate (not that of the conducting sphere) is operational for the non-directional screening. So we defined the polarizability upper bound of atom A as

$$\alpha_A^{\text{upperbound}} = \frac{V_A}{2\pi} = \frac{1}{2\pi} \oint \frac{\rho_A(\vec{r}_A)}{\sum_{B,L} \rho_B(\vec{r}_B)} d^3\vec{r}_A \quad (49)$$

where V_A is the volume of atom A in the material.

During MCLF calculation, if the calculated non-directionally screened polarizability is higher than this upper bound, the result will be replaced by the upper bound polarizability. This procedure is carried out by a smooth min function

$$\alpha_A = \text{smooth_min}(\alpha_A^{\text{est}}, \alpha_A^{\text{upperbound}}) \quad (50)$$

$$\text{smooth_min}(a, b) = \frac{\sqrt{a * b}}{\left(\left(\frac{a}{b} \right)^\gamma + \left(\frac{b}{a} \right)^\gamma - 1 \right)^{\frac{1}{2\gamma}}} \quad (51)$$

where $\gamma = 25$ because this value provides a smooth curve with less than 0.58% deviation from $\max(\min(a, b), 0)$. If either a or b is ≤ 0 , the function is set to return a value of 0. Using a smooth curve, instead of simply $\min(a, b)$, is required ensure the forces (i.e., energy derivatives with respect to atom displacements) are continuous functions.

4.3 M scaling to describe both surface and buried atoms

Section 3.2 above shows the polarizability of an isolated atom scales approximately proportional to its effective radius to the 4th power. Brinck et al. showed the polarizability of a polyatomic molecule is approximately proportional to its volume divided by an effective electron removal energy.⁸³ As shown in Section 4.2, the polarizability of a conducting plate or sphere is proportional to its volume. For non-conducting fluids, the Clausius-Mossotti relationship describes a polarizability-to-volume ratio that is a weak function of the dielectric constant.⁶⁴ This implies the polarizability-to-volume ratio of a buried atom is approximately proportional to its volume. This means the atom-in-material polarizability transitions from approximately 4th power to 3rd power dependence on the atom's effective radius as the atom goes from isolated to buried.

We define m to quantify how exposed an atom is:

$$m = \frac{\langle r^4 \rangle_{\text{weighted}}}{\langle r^4 \rangle} \quad (52)$$

$$\langle r^4 \rangle_{\text{weighted}} = \oint \rho_A(\vec{r}_A) * (r_A)^4 * \frac{\left(\left\langle \left(\frac{w_A(r_A)}{W(\vec{r})} \right)^4 \right\rangle_{r_A} \right)}{\left(\left\langle \left(\frac{w_A(r_A)}{W(\vec{r})} \right)^3 \right\rangle_{r_A} \right)} d^3\vec{r}_A \quad (53)$$

where $\rho_A(\vec{r}_A)$ is the electron density of atom A, $w_A(\vec{r}_A)/W(\vec{r})$ is the fraction of the total electron density $\rho(\vec{r})$ at position \vec{r} that is assigned to atom A, and $\langle \rangle_{r_A}$ means the spherical average. m equals 1 for an isolated atom and goes towards zero when an atom gets more and more buried.

The modified unscreened scaling law for α now has the form

$$\Omega = \alpha^{\text{ref}} \left(\frac{\langle r^3 \rangle^{\text{ref}}}{\langle r^3 \rangle^{\text{AIM}}} \right)^{3.1657} \left(\frac{\langle r^4 \rangle^{\text{AIM}}}{\langle r^4 \rangle^{\text{ref}}} \right)^{3.3372} \left(\frac{n^{\text{AIM}}}{n^{\text{ref}}} \right)^{0.2892} \quad (54)$$

$$\alpha^{\text{unscreened}} = \left(C \langle r^3 \rangle \right)^{(1-m)} \Omega^m \quad (55)$$

When $m = 1$, it becomes the isolated atom scaling law in Section 3. When $m = 0$, it becomes

$$\lim_{m \rightarrow 0} \alpha^{\text{unscreened}} = C \langle r^3 \rangle \quad (56)$$

There are two primary justifications for Eq. (56). First, the Clausius-Mossotti relation and conduction limit upper bound show the polarizability of a condensed phase (e.g., liquid or solid) is approximately proportional to its volume. Here, $\langle r^3 \rangle$ is a sort of proxy for the volume of an atom, so the sum of $\langle r^3 \rangle$ moments for atoms in a material is a sort of proxy for the material's volume. The Clausius-Mossotti relation and conduction limit upper bound show the polarizability-to-volume ratio of a material is not strongly element-dependent and has modest dependence on the material's dielectric constant (see Eqs. (106)–(107)). Second, in condensed phases electrons undergo chemical potential equilibration that transfers some electron density from the least electronegative elements to the most electronegative elements. Accordingly, the chemical potential of the equilibrated condensed phase is not as extreme as either the most electropositive elements nor the most electronegative elements. For atoms in condensed materials, the polarizability to $\langle r^3 \rangle$ moment ratios will typically be lower than an isolated alkali metal atom (extremely electropositive) on the one hand and higher than an isolated fluorine atom (extremely electronegative) on the other hand. The polarizability to $\langle r^3 \rangle$ moment ratios of atoms in condensed materials thus exhibits a narrower range of values than for isolated atoms. Group 14 elements have approximately equal tendency to gain or lose electrons, thereby readily forming both positive and negative oxidation states. The isolated atom polarizability to $\langle r^3 \rangle$ moment ratios of Group 14 elements are approximately independent of the periodic row: C (0.34), Si (0.37), Ge (0.34), Sn (0.32), and Pb (0.31). Accordingly, we expect the same constant C in Eq. (56) will work for both light and heavy elements.

Tests on the polarizabilities of 28 solids were performed to optimize C . The criteria are that without any upper bound imposed, the MRE should be 0 – 10% and after the upper bound is imposed, the MARE should be $\leq \sim 10\%$. These criteria make sure the scaling law is as accurate as possible without the upper bound and the upper bound will not decrease the accuracy. Table 4 shows that $C = 0.4$ is optimal. This value is slightly higher than the polarizability to $\langle r^3 \rangle$ moment ratios of isolated Group 14 atoms.

Table 4: Comparison of the % error in the polarizability of 28 solids as a function of C value. “NU” means no upper bound is applied and “U” means the upper bound is applied.

	C = 0.35 NU	C = 0.4 NU	C = 0.45 NU	C = 0.35 U	C = 0.4 U
MRE	1.14%	6.87%	11.91%	-12.22%	-9.14%
MARE	26.75%	27.26%	29.09%	13.59%	11.89%

The characteristic frequency ω_p also has different expressions for isolated and buried atoms. For an isolated atom (i.e., $m = 1$), ω_p should be proportional to $\alpha^{-1/2}$ which has been captured by the isolated atom scaling law. For a buried atom (i.e., $m \ll 1$), the polarizability becomes nearly proportional to $\langle r^3 \rangle^{\text{AIM}}$ and C_6 remains proportional to the atom's effective radius to the sixth power; therefore, ω_p is less sensitive to changes in α when the atom is buried. So the scaling for ω_p has been modified to:

$$\gamma = \left(\frac{N^{\text{ref}}}{N^{\text{AIM}}} \right)^{0.3167} \left(\frac{\langle r^3 \rangle^{\text{AIM}}}{\langle r^3 \rangle^{\text{ref}}} \right)^{3.7003} \left(\frac{\langle r^4 \rangle^{\text{ref}}}{\langle r^4 \rangle^{\text{AIM}}} \right)^{3.2228} \quad (57)$$

$$\zeta = \frac{\Omega}{C \langle r^3 \rangle^{\text{AIM}}} \quad (58)$$

$$\omega_p = \zeta^{(1-m^2)/4} \gamma \omega_p^{\text{ref}} \quad (59)$$

When $m = 1$, Eq. (59) reduces to the isolated atom scaling law in Section 3. A complete explanation and derivation of Eq. (59) is given in Section S3 of ESI†. Figure 9 shows pseudocode for the input file reading and unscreened calculation using m-scaling.

```

Begin input file reading
  Read the following quantities needed to perform the unscreened scaling calculations: Natoms; atomic
  numbers {ElementA} or element symbols; net atomic charges {NA}, Rcubed, Rfourth, and
  weighted_Rfourth atom-in-material radial moments
  Read the following quantities needed to perform the screening calculations: periodic lattice vectors (if any),
  (X, Y, Z) coordinates of each atom, atom-in-material volumes {VA}
  If element symbols were read, then convert to atomic numbers {ElementA}
  Calculate the polarizability upper bound for every atom:  $\alpha_A^{\text{upper\_bound}} = V_A / (2\pi)$ 
  Calculate the number of electrons for every atom:  $N_A = \text{Element}_A - q_A$ 
End input file reading

Begin unscreened scaling
  Loop over atoms in the unit cell A = 1, 2, ... Natoms
    Calculate m = weighted_Rfourth_moment(A) / Rfourth_moment(A)
    Calculate  $\alpha^{\text{unscreened}}(A)$  using Eq. 55
    Calculate  $\omega_p(A)$  using Eq. 59
    Calculate unscreened_C6(A) =  $0.75 (\alpha^{\text{unscreened}}(A))^2 \omega_p(A)$ 
    Calculate unscreened_Rdamp(A) = Rdamp_ref(Element(A))  $\left( \frac{\text{unscreened\_C6}(A)}{C6\_ref(\text{Element}(A))} \right)^{1/9}$ 
  End loop
  Initialize total_unscreened_polarizability = 0 and total_unscreened_C6 = 0
  Loop over atoms in the unit cell A = 1, 2, ... Natoms
    Loop over atoms in the unit cell B = 1, 2, ... Natoms
      total_unscreened_C6 = total_unscreened_C6
      +  $\frac{2(\text{unscreened\_C6}(A))(\text{unscreened\_C6}(B))(\alpha^{\text{unscreened}}(A))(\alpha^{\text{unscreened}}(B))}{(\alpha^{\text{unscreened}}(B))^2 \text{unscreened\_C6}(A) + (\alpha^{\text{unscreened}}(A))^2 \text{unscreened\_C6}(B)}$ 
    End loop
    total_unscreened_polarizability = total_unscreened_polarizability +  $\alpha^{\text{unscreened}}(A)$ 
  End loop
End unscreened scaling

```

Figure 9: Pseudocode for the input file reading and unscreened calculation using m-scaling.

4.4 Iterative polarizability screening

The non-directional, fluctuating, and static dipole interaction tensors are defined as:

$$\tau_{\text{non-dir}}^{\text{AB}}(\mathbf{u}) = \sum_L \frac{4}{3\sqrt{\pi}} \frac{f_{\text{cutoff}}(\mathbf{r}^{\text{AB},L})}{(\sigma_{\text{AB}}(\mathbf{u}))^3} e^{-\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})}\right)^2} \quad (60)$$

$$\tau_{\alpha\beta \text{ fluctuating}}^{\text{AB}}(\mathbf{u}) = \sum_L f_{\text{cutoff}}(\mathbf{r}^{\text{AB},L}) \cdot f_{\text{MBS}}(\mathbf{r}^{\text{AB},L}) \cdot \left[\frac{(\mathbf{r}^{\text{AB},L})^2 \delta_{\alpha\beta} - 3r_{\alpha}^{\text{AB},L} r_{\beta}^{\text{AB},L}}{(\mathbf{r}^{\text{AB},L})^5} \left(\text{erf}\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})}\right) - \frac{2}{\sqrt{\pi}} \frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})} e^{-\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})}\right)^2} \right) + \frac{4}{3\sqrt{\pi}} \frac{1}{(\sigma_{\text{AB}}(\mathbf{u}))^3} \frac{3r_{\alpha}^{\text{AB},L} r_{\beta}^{\text{AB},L} - (\mathbf{r}^{\text{AB},L})^2 \delta_{\alpha\beta}}{(\mathbf{r}^{\text{AB},L})^2} e^{-\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})}\right)^2} \right] \quad (61)$$

$$\tau_{\alpha\beta \text{ static}}^{\text{AB}} = \sum_L f_{\text{cutoff}}(\mathbf{r}^{\text{AB},L}) \cdot \left[\frac{(\mathbf{r}^{\text{AB},L})^2 \delta_{\alpha\beta} - 3r_{\alpha}^{\text{AB},L} r_{\beta}^{\text{AB},L}}{(\mathbf{r}^{\text{AB},L})^5} \left(\text{erf}\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}}\right) - \frac{2}{\sqrt{\pi}} \frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}(\mathbf{u})} e^{-\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}}\right)^2} \right) + \frac{4}{3\sqrt{\pi}} \frac{1}{(\sigma_{\text{AB}})^3} \frac{3r_{\alpha}^{\text{AB},L} r_{\beta}^{\text{AB},L} - (\mathbf{r}^{\text{AB},L})^2 \delta_{\alpha\beta}}{(\mathbf{r}^{\text{AB},L})^2} e^{-\left(\frac{\mathbf{r}^{\text{AB},L}}{\sigma_{\text{AB}}}\right)^2} \right] \quad (62)$$

The subscripts α and β refer to spatial indices x, y, and z. This enables the non-directional and directional components to be separately screened. As described in Section 4.2, the conduction limit upper bound applies to the non-directionally screened (not the directionally screened) components. Computing static polarizabilities (i.e., system response to constant externally applied electric field) requires directionally screened polarizabilities. As discussed in Section 4.6, parameterizing polarizable force fields requires non-directionally screened polarizabilities.

$f_{\text{cutoff}}(\mathbf{d}_{\text{Ab}})$ is a smooth cutoff function that smoothly turns off dipole-dipole interactions between atoms as the distance between them increases:

$$f_{\text{cutoff}}(\mathbf{d}_{\text{Ab}}) = \left[1 - \exp\left(-20\left(1 - \frac{\mathbf{d}_{\text{Ab}}}{\mathbf{d}_{\text{cutoff}}}\right)^3\right) \right] \cdot (1 - H(\mathbf{d}_{\text{Ab}} - \mathbf{d}_{\text{cutoff}})) \quad (63)$$

where $\mathbf{d}_{\text{Ab}} = \mathbf{r}^{\text{AB},L}$ is the distance between atom A and the image b in bohr. $\mathbf{d}_{\text{cutoff}}$ is the dipole interaction cutoff length. H is the Heaviside step function. As shown in Fig. S1 of ESI†, this function smoothly decreases from ~ 1 at $\mathbf{d}_{\text{Ab}} = 0$ to zero when $\mathbf{d}_{\text{Ab}} \geq \mathbf{d}_{\text{cutoff}}$. The power of three in the exponential ensures that both the first and second derivatives are continuous at $\mathbf{d}_{\text{Ab}} = \mathbf{d}_{\text{cutoff}}$, which is required for frequency calculations. The factor of 20 in the exponent provides a balanced cutoff function steepness. Specifically, $f_{\text{cutoff}}(\mathbf{d}_{\text{Ab}} \leq 0.5\mathbf{d}_{\text{cutoff}}) \geq 0.9179$ ensuring that all positions within half the cutoff distance are counted at nearly full strength.

Imagine a parameter $0 \leq \lambda \leq 1$ that continuously turns on a particular screening type (e.g., non-directional, fluctuating, or static). When $\lambda = 0$, the corresponding screening type is fully off. When $\lambda = 1$, the corresponding screening type is fully on. Thus, the corresponding screening process can be envisioned as transitioning continuously from $\lambda = 0$ to $\lambda = 1$. We expect the partially screened system

(i.e. $0 < \lambda < 1$) to have a polarizability intermediate between that for $\lambda = 0$ and $\lambda = 1$. Since the Gaussian width $\sigma_{AB}(u)$ depends on the polarizability (eqn (19)), the MCLF method continuously updates $\sigma_{AB}(u)$ during the screening process. In contrast, the TS-SCS method uses $\sigma_{AB}(u)$ corresponding to $\lambda = 0$ for the entire screening process, which does not allow the Gaussian width to equilibrate during the screening process.

The screening of each tensor is now an iterative process. In each iteration, we compute

$$Q_{j+1} = D_j + \Delta_j \tau_j \quad (64)$$

where Δ_j is the screening increment. Then P_{j+1} is computed as the inverse of Q_{j+1} . The screening process is divided into non-directional and directional screening. Directional screening has two separate parts: screening for the static polarizability and screening for the fluctuating polarizabilities. These parts are summarized below.

Non-directional screening: For non-directional screening, both D and τ are N_{atoms} by N_{atoms} matrices. In the first iteration, $j = 1$ and D is constructed using $(\alpha^{\text{unscreened}}(u))^{-1}$ for the respective atoms along the diagonal and zeros elsewhere, and $\sigma_{AB}(u)$ is calculated from $\alpha^{\text{unscreened}}(u)$. τ_j is defined in eq. (60) where $\sigma_{AB}(u)$ is computed using $P_{A,j}$ and $P_{B,j}$ in each iteration j ,

$$P_{A,j}^{\text{non-dir}}(u) = \sum_B P_j^{AB}(u) \left(\frac{2\alpha_A^{\text{unscreened}}(u)}{\alpha_A^{\text{unscreened}}(u) + \alpha_B^{\text{unscreened}}(u)} \right) \quad (65)$$

D_j is the matrix which has $(P_{A,j}^{\text{non-dir}}(u))^{-1}$ of each atom on the diagonal and the rest is zero. The new D_j and τ_j are fed into eq. (64) for the next iteration. The calculation continues until

$$\sum_j \Delta_j = 1 \quad (66)$$

On the last iteration, $\alpha_A^{\text{raw_non-dir}}(u)$ is calculated via Eq. (44).

Because the polarizability upper bound must be calculated for uniform applied electric field and does not include any directional interactions between dipoles, it is applied at the end of non-directional screening and before directional screening. For each u value, the polarizability upper bound was applied to the raw non-directionally screened polarizability (i.e., $\alpha_A^{\text{raw_non-dir}}(u)$) via the following equation to generate

$$\alpha_A^{\text{non-dir}}(u) = \text{smooth_min} \left(\alpha_A^{\text{raw_non-dir}}(u), \frac{\alpha_A^{\text{upperbound}}}{1 + \left(\frac{\omega(u)}{w p_A} \right)^2} \right) \quad (67)$$

Note that

$$\alpha_A^{\text{force-field}} = \alpha_A^{\text{non-dir}}(u = \text{Nimfreqs}) \quad (68)$$

Directional screening for static polarizability: This procedure is only done at zero frequency. For directional screening, D and τ (eq. (62)) are $3 \times N_{\text{atoms}}$ by $3 \times N_{\text{atoms}}$ matrixes. D is a block diagonal matrix which has the inverse polarizability tensor of each atom on the block diagonal and the rest is zero. In the

first iteration, D and σ_{AB} are obtained using $\alpha_A^{\text{force-field}}$. P_j undergoes a partitioned partial contraction (eqn (46)) to obtain partially screened polarizability tensors for each atom. The isotropic polarizability of each atom is 1/3 of the trace of their tensors. From this partially screened polarizability, the associated Gaussian screening width (eq. (19)) is obtained and fed into τ_j . The polarizability tensor of each atom is inversed to construct a new D matrix. The new D_j and τ_j are fed into eq. (64) for the next iteration. The calculation continues until Δ_j sums to 1.

Directional screening for fluctuating polarizabilities: D and τ (eqn (61)) are $3*N_{\text{atoms}}$ by $3*N_{\text{atoms}}$ matrixes. D is a block diagonal matrix which has the inverse polarizability tensor of each atom on the block diagonal and the rest is zero. In the first iteration, D and $\sigma_{AB}(u)$ are calculated using $\alpha_A^{\text{non-dir}}(u)$. P_j undergoes a partitioned partial contraction (eq. (45)) to obtain partially screened polarizability tensors for each atom. The isotropic polarizability of each atom is 1/3 of the trace of their tensors. From this partially screened polarizability, the associated Gaussian screening width (eq. (19)) is obtained and fed into τ_j . The polarizability tensor of each atom is inversed to construct a new D matrix. The new D_j and τ_j are fed into eq. (64) for the next iteration. The calculation continues until Δ_j sums to 1. On the last iteration, $\alpha_A^{\text{screened}}(u)$ is calculated via Eqn (45). This procedure is done at multiple frequencies and the resulting $\{\alpha_A^{\text{screened}}(u)\}$ are fed into the Casimir-Polder integral to calculate the C_6 dispersion coefficient of each atom. Note that

$$\alpha_A^{\text{low-freq}} = \alpha_A^{\text{screened}}(u = N_{\text{mfreqs}}) \quad (69)$$

4.5 Multibody screening parameter (MBSP)

When a uniform external electric field is applied, the atomic dipoles induced by the field will align and atoms will interact with not only their neighbors but also atoms far away. The dispersion force, however, is caused by fluctuating dipoles. The fluctuating dipoles of the atoms will align with their neighbors but out of sync with atoms far away. The MBSP controls the length scale over which directional alignment persists:

$$f_{\text{MBS}}(d_{Ab}) = \exp\left(-\frac{d_{Ab}}{(r_{\text{damp},A}^{\text{unscreened}} + r_{\text{damp},B}^{\text{unscreened}}) * \text{MBSP}}\right) \quad (70)$$

First-order exponential decay (Eq. (70)) is the natural choice for the directional alignment function, because if 50% of the directional alignment persists over a distance d_{half} , then over a distance $2*d_{\text{half}}$ the expected persistence of directional alignment will be $(50\%)*(50\%)=25\%$. As defined in the ESI†, $r_{\text{damp},A}^{\text{unscreened}}$

and $r_{\text{damp},B}^{\text{unscreened}}$ are unscreened damping radii of the atoms-in-material. We optimized MBSP with the C_6 of the 12 polyacenes and 6 fullerenes, the same set studied in Section 6.5. Table 5 shows the MRE and MARE of different MBSP values of which 2.5 is optimal.

Table 5: Comparison of the % error in C_6 of polyacenes and fullerenes using different MBSP values.

MBSP →		2.0	2.25	2.5	2.75
MRE	Polyacenes	-7.37%	-4.73%	-2.38%	-0.28%
	Fullerenes	5.16%	6.09%	6.84%	7.45%
MARE	Polyacenes	10.76%	9.14%	7.77%	7.01%
	Fullerenes	5.16%	6.09%	6.84%	7.45%

4.6 Flow diagram and explanation for three polarizability types

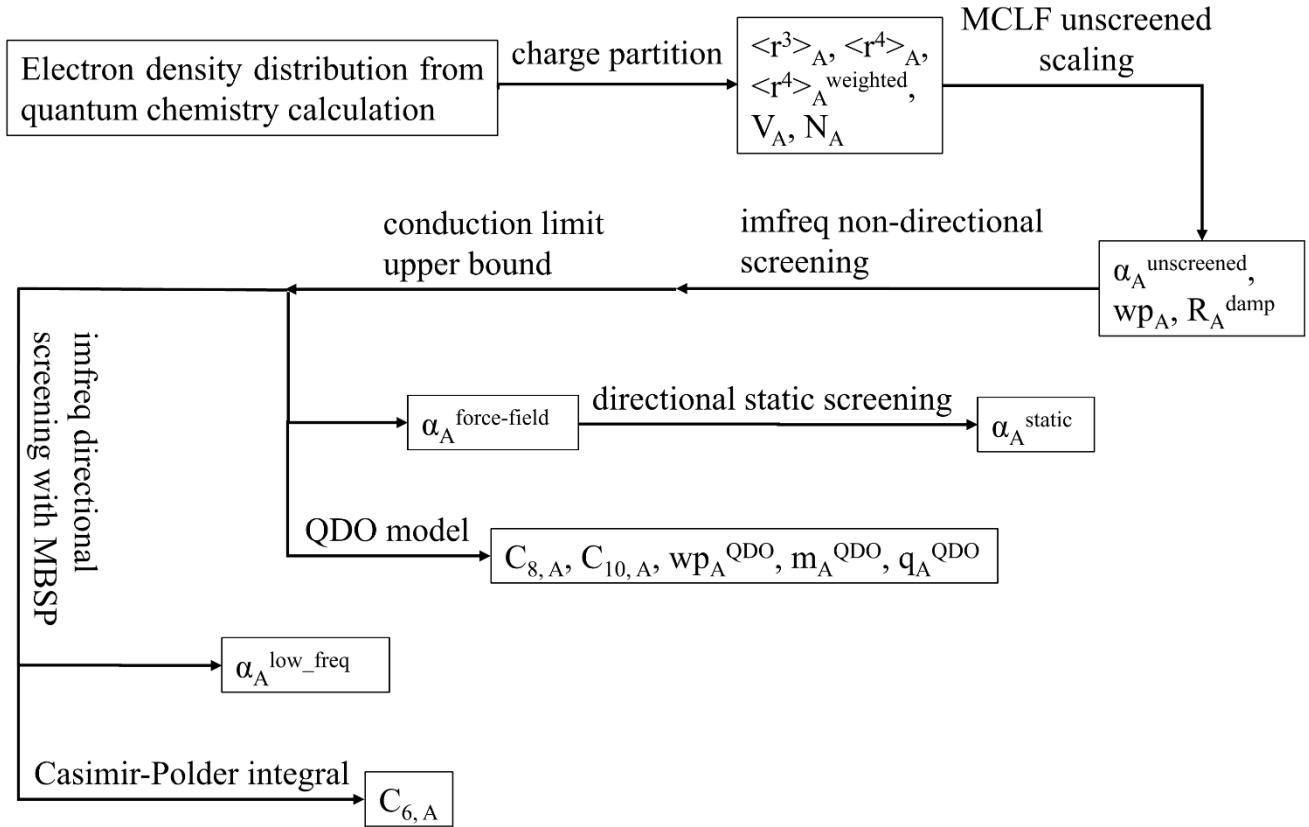


Figure 10: Flow diagram for MCLF method

Figure 10 is the flow diagram for MCLF method. This method yields three kinds of polarizabilities: $\alpha^{\text{force-field}}$, α^{static} and $\alpha^{\text{low_freq}}$. $\alpha^{\text{force-field}}$ is the polarizability with no directional ordering of the electric field and mainly used in force-field simulations. α^{static} is the static polarizability in a constant applied electric field. $\alpha^{\text{low_freq}}$ describes the short-range ordering and long-range disordering of the fluctuating dipoles present in the dispersion interaction.

Because polarizability is a multibody effect, the polarizability of a molecule is not the sum of polarizabilities of its isolated constituent atoms. Directional interactions between atoms creates components to the molecular polarizability that do not exist for the isolated atoms.^{26, 27} Directional dipole-

dipole interactions^{26, 27} are considered during classical atomistic (e.g., molecular dynamics or Monte Carlo) simulations that utilize polarizable force fields.^{14, 84} To avoid double counting these directional dipole-dipole interactions, the force field's atomic polarizability parameters must be the non-directionally screened values.⁸⁵⁻⁸⁸ Many authors proposed direct additive partitioning of the quantum-mechanically computed molecular polarizability tensor into constituent atoms.^{8, 89-94} Because those values already incorporate directional dipole-dipole interactions, their use in force fields would result in double counting the directional dipole-dipole interactions; therefore, we do not recommend their use as force field parameters. Our method provides both the non-directionally screened and directionally screened atomic polarizabilities, and the non-directionally screened values should be used as polarizable force-field parameters. The directional screening then arises during the course of the classical atomistic simulation.

5. Computational algorithm and timing results

5.1 Lists of interacting atom pairs

Two separate lists of interacting atom pairs are prepared after the unscreened calculation and before the screened calculation. For convenience, we refer to these as the ‘small’ and ‘large’ lists, where ‘small’ and ‘large’ refer to the interaction cutoff distance. Without loss of generality, the first atom in each pair can be considered to reside within the reference unit cell. The second atom is located somewhere within the dipole interaction cutoff length of the first atom. Therefore, constructing these lists involves two nested loops over atoms: (i) one loop over all atoms in the reference unit cell and (ii) one loop over enough images of the reference unit cell to completely enclose a sphere of dipole interaction cutoff radius around every atom in the reference unit cell. Only translation symmetry unique atom pairs need to be included in the loops over atom pairs. As explained in the earlier bond order article, an atom image pair is translation symmetry unique iff. at least one of the following criteria is satisfied: (i) the index numbers of the atoms are different (e.g., atom 210 and atom 1056), (ii) $L_1 > 0$, (iii) $L_1 = 0$ and $L_2 > 0$, or (iv) $L_1 = L_2 = 0$ and $L_3 > 0$.⁵⁸

The small lists all atom pairs having ‘overlapping’ Gaussian dipole model densities as defined by the cutoff criterion

$$v_{Ab}^{\text{unscreened}} = \frac{d_{Ab}}{\sigma_{AB}^{\text{unscreened}} (u = \text{Nimfreqs})} \leq v_{\text{cutoff}} \quad (71)$$

This criteria ensures $\text{erfc}(v_{Ab}) \approx \exp(-v_{Ab}^2) \approx 0$ for any atom pair not included in the list. The value $v_{\text{cutoff}} = 5^{4/3}$ was chosen such that even if the (partially) screened polarizability is larger than the unscreened polarizability by up to a factor of five, then $\text{erfc}(v_{Ab}^{\text{screened}}) \leq \text{erfc}(5) = 1.5 \times 10^{-12}$ and $\exp(-v_{Ab}^{\text{screened}}^2) \leq \exp(-25) = 1.4 \times 10^{-11}$. The following information was saved in the small list array for each included atom pair: the index numbers of the first and second atoms, the translation integers for the second atom, the atomic number of each atom, d_{Ab} , the value of the smooth cutoff function, the value of the multibody screening function times the smooth cutoff function, and the following six tensor components: $\eta_{1,1} = 3(\Delta x)^2/d_{Ab}^2 - 1$, $\eta_{1,2} = \eta_{2,1} = 3\Delta x \Delta y / d_{Ab}^2$, $\eta_{1,3} = \eta_{3,1} = 3\Delta x \Delta z / d_{Ab}^2$, $\eta_{22} = 3(\Delta y)^2/d_{Ab}^2 - 1$, $\eta_{2,3} = \eta_{3,2} = 3\Delta y \Delta z / d_{Ab}^2$, $\eta_{3,3} = 3(\Delta z)^2/d_{Ab}^2 - 1$. Here, Δx , Δy , and Δz are the Cartesian components of $\vec{R}_A - \vec{R}_B$.

The large lists all atom pairs (A, B) in the reference unit cell for which atom A (first atom) is located a distance less than dipole interaction cutoff length to at least one image of atom B (second atom). The self-pair (A, A) is included iff. one of the non-trivially translated images of atom A is located a distance less than dipole interaction cutoff length to atom A in the reference unit cell. The number of atom pairs in the large list is always $\leq N(N+1)/2$, where N is the number of atoms in the unit cell. For example, a NaCl crystal containing two atoms in the reference unit cell would contain three pairs in the large list: Na-Na, Na-Cl, and Cl-Cl. For each pair (A, B) in the large list, a loop is performed over all atom b images located within dipole interaction cutoff length of atom A, and the following sums are accumulated and stored:

$$\vec{\vec{g}}^{AB} = \sum_L f_{\text{cutoff}}(\mathbf{d}_{Ab}) \vec{\vec{n}}^{Ab} \quad (72)$$

$$\vec{\vec{h}}^{AB} = \sum_L f_{\text{cutoff}}(\mathbf{d}_{Ab}) f_{\text{MBS}}(\mathbf{d}_{Ab}) \vec{\vec{n}}^{Ab} \quad (73)$$

Since $\vec{\vec{g}}$ and $\vec{\vec{h}}$ are symmetric with respect to exchange of the spatial indices, only six components need to be computed and stored for each.

Having separate small and large lists provides the following computational efficiencies. First, non-directional screening only needs to be performed over the ‘overlapping’ atoms contained in the small list. Second, the computationally expensive erfc function only needs to be evaluated for atom pairs in the small list. This provides computational savings during directional screening that first loops over all pairs in the small list and then over all pairs in the large list. Third, the large list contains pre-computed sums over all periodic images of interacting atom pairs. For non-overlapping atoms, this avoids re-computing any sums over periodic images at each screening increment and each frequency point.

These two lists of interacting atom pairs are used as follows. For each screening increment of each frequency point, non-directional screening loops over all atom pairs in the small list to compile the necessary dipole-dipole interaction terms. For each screening increment of each frequency point, directional screening first loops over all atom pairs in the small list and then over all atom pairs in the large list to compile the associated dipole-dipole interaction terms. More details are provided in Section 5.3 below.

Table 6: Effect of dipole interaction cutoff length on the MCLF computed properties of graphene.

	10 bohr	25 bohr	50 bohr	75 bohr	100 bohr
C_6	37.91	50.33	52.43	52.31	52.31
α^{static}	11.42	18.13	20.15	19.95	19.96
$\alpha^{\text{low_freq}}$	9.75	11.81	12.14	12.12	12.12
$\alpha^{\text{force-field}}$	7.03	7.03	7.03	7.03	7.03

Table 6 studies the effect of dipole interaction cutoff length on computed precision. Graphene was chosen as a test system, because it has strong long-range dipole-dipole coupling. As shown in Table 6, results differed by <1% between 50 and 100 bohr cutoffs. α^{static} differed by ~10% between 25 and 50 bohr cutoffs. A 10 bohr cutoff was woefully inadequate. For dense materials with large unit cells, doubling the dipole interaction cutoff length increases the computational cost by approximately eight-fold. We selected dipole interaction cutoff length = 50 bohr as a good compromise between computational cost and precision. This value was used for all other results in this article.

5.2 Integration over imaginary frequencies

The Casimir-Polder integral relates the C_6^{AB} dispersion coefficient between two subsystems A and B to the product of their fluctuating polarizabilities at imfreq ω integrated over all imfreqs:⁶⁷

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty \alpha_A(u) \alpha_B(u) du \quad (74)$$

As shown in Eq. (74), the integration limits are zero to infinity. For convenience, we used the substitution of variables in eqn (15) to make both integration limits finite. Differentiating eqn (15) yields

$$d\omega = -\frac{\text{Nimfreqs}}{u^2} du \quad (75)$$

which upon substitution into eqn (74) for A=B gives

$$C_6^{AA} = \frac{3}{\pi} \int_0^{\text{Nimfreqs}} \frac{(\alpha_A(u))^2}{u^2} du \quad (76)$$

$u=0$ corresponds to infinite imfreq. Near infinite imfreq, the polarizability becomes inversely proportional to ω^2 (e.g., eqn (20)).⁹⁵ At this limit, the integrand in Eq. (76) simplifies to

$$\lim_{u \rightarrow 0} \frac{(\alpha_A(u))^2}{u^2} \text{Nimfreqs} \approx \lim_{u \rightarrow 0} \frac{(\text{cons tan } t / \omega^2)^2 \text{Nimfreqs}}{u^2} \quad (77)$$

Substituting eqn (15) into (77) gives

$$\lim_{u \rightarrow 0} \frac{(\text{cons tan } t)^2 \text{Nimfreqs}}{u^2 \left(\frac{\text{Nimfreqs}}{u} - 1 \right)^4} = \lim_{u \rightarrow 0} \frac{(\text{cons tan } t)^2 \text{Nimfreqs}}{\left(\frac{\text{Nimfreqs}}{u} - 1 \right)^2 (\text{Nimfreqs} - u)^2} = 0 \quad (78)$$

Therefore, the $u=0$ point contributes nothing to the integral.

We numerically integrated using Richardson extrapolation (i.e., Romberg integration).⁹⁶⁻⁹⁸ Dividing the (0,1] interval into 2^G segments and performing Romberg integration of order (G, G) yields an integration error of the order $2^{-G(2G+2)}$.⁹⁶ Normally, Romberg integration of 2^G segments corresponds to 2^{G+1} integration points. Since the $u=0$ point contributes nothing to the integral, this leaves only $\text{Nimfreqs} = 2^G$ nontrivial integration points. The ESI† contains the Romberg integration weights $c_{G,u}^{\text{Romberg}}$ of these 2^G integration points for $G = 1$ to 5. The integral is computed using the following sum:

$$C_6^{AA} = \sum_{u=1}^{\text{Nimfreqs}} c_{G,u}^{\text{Romberg}} \left(\frac{\text{Nimfreqs} \cdot \alpha_A(u)}{u} \right)^2 \quad (79)$$

where

$$\sum_{u=0}^{\text{Nimfreqs}} c_{G,u}^{\text{Romberg}} = 1 \quad (80)$$

Table 7 shows computed AIM C_6 coefficients for four materials. The K bcc solid showed a small difference (0.3%) in the C_6 coefficient between 16 and 32 integration points. All other results were virtually identical for 16 and 32 integration points. Moreover, all results for 8 integration points differed by <1% from the higher integration points. This shows the results are highly converged for 16 integration points, which is the value we chose.

Table 7: Effect of Rhomberg integration order (G) and number of integration points (Nimfreqs=2^G) on the computed atom-in-material C₆ coefficients in atomic units.

G →	3	4	5
Nimfreqs →	8	16	32
graphene	52.39	52.43	52.43
K bcc solid	444.29	448.35	447.00
NaF solid	11.69 (Na) 45.13 (F)	11.69 (Na) 45.26 (F)	11.69 (Na) 45.26 (F)
C60	29.57	29.67	29.67

Note: The Rhomberg integration order is set such that $Nimfreqs = 2^{Rhom_order}$.

Begin screened dispersion coefficients and QDO parameters calculations

Initialize screened_C6(A)=0 and nondir_C6(A)=0

Loop over the imfreq points u = 1, 2, ... Nimfreqs

Calculate temp = (Nimfreqs / u)² (3/π) Rhom_coeff (Rhom_order, u)

Loop over atoms in the unit cell A = 1, 2, ... Natoms

Calculate: screened_C6(A) = screened_C6(A) + (α^{screened}(A, u))² temp

Calculate: nondir_C6(A) = nondir_C6(A) + (α^{non-dir}(A, u))² temp

End loop

End loop

Loop over atoms in the unit cell A = 1, 2, ... Natoms

Calculate C8(A) = e^{1.7327} (nondir_C6(A))^{3.8305/3} (Rfourth_moment(A) / Rcubed_moment(A))^{0.339}

Calculate C10(A) = (49/40) (C8(A))² / nondir_C6(A)

Calculate wp_QDO(A) = 4nondir_C6(A) / 3α^{force-field}(A)

Calculate m_QDO(A) = 15α^{force-field}(A) / 4C8(A)

Calculate q_QDO(A) = -√(20(nondir_C6(A))² / 3C8(A))

End loop

Initialize total_screened_C6 = 0

Loop over atoms in the unit cell A = 1, 2, ... Natoms

Loop over atoms in the unit cell B = 1, 2, ... Natoms

total_screened_C6 = total_screened_C6

+ 2(screened_C6(A))(screened_C6(B))(α^{low_freq}(A))(α^{low_freq}(B))
+ (α^{low_freq}(B))² screened_C6(A) + (α^{low_freq}(A))² screened_C6(B)

End loop

End loop

End screened dispersion coefficients and QDO parameters calculations

Figure 11: Pseudocode for computing the screened dispersion coefficients and QDO parameters.

Although the mixed C₆^{AB} dispersion coefficients could in principle be computed from the Casimir-Polder integral using a similar procedure, this would involve many integrations for materials containing

thousands of atoms in the unit cell. Therefore, we used the following mixing formula which is consistent with both Padé approximation⁷¹ and QDO⁷ models:

$$C_6^{AB} = \frac{2\alpha_A^{\text{low_freq}}\alpha_B^{\text{low_freq}}C_6^{AA}C_6^{BB}}{(\alpha_B^{\text{low_freq}})^2C_6^{AA} + (\alpha_A^{\text{low_freq}})^2C_6^{BB}} \quad (81)$$

In our method, the polarizabilities appearing in Eq. (81) must be $\alpha^{\text{low_freq}}$, because these are the polarizabilities associated with the dispersion interaction. Of note, the TS and TS-SCS methods use similar mixing formula, except the polarizabilities appearing in the mixing formula are α^{TS} and $\alpha^{\text{TS-SCS}}$,^{10, 38} because those methods do not yield $\alpha^{\text{low_freq}}$. Figure 11 illustrates a pseudocode for computing the screened dispersion coefficients and QDO parameters.

5.3 Avoiding direct inversion of large matrices

Direct inversion of large matrices is extremely computationally expensive. Gaussian elimination with partial pivoting (GEPP) is a common matrix inversion algorithm that exhibits numerical instability for some matrices but is usually stable in practice.^{99, 100} GEPP, QR factorization, Cholesky and LDL decomposition (for positive definite Hermitian matrices), and other common matrix inversion algorithms have computational costs scaling proportional to the number of rows (Nrows) cubed.^{100, 101} The more complicated Strassen algorithm has scaling proportional to $\text{Nrows}^{\log(7)/\log(2)} = \text{Nrows}^{2.807...}$.¹⁰¹

A direct matrix inversion algorithm would first construct the dipole interaction tensor, then invert it to get the multibody polarizability matrix, and then contract the multibody polarizability matrix to get the atom-in-material polarizability tensors.²⁶ Consider a material containing 250000 atoms in the unit cell. For this material, the dipole interaction tensor has 750000 rows and an equal number of columns. The computational cost of the matrix inversion step would be on the order of $\text{Nrows}^3 = (750000)^3 = \sim 4 \times 10^{17}$ floating point operations. This corresponds to ~ 117 computational hours on a teraflop computer. Since the dipole interaction tensor would need to be inverted at each imfreq integration point, a TS-SCS or MCLF calculation on this material would take thousands of computational hours when using direct matrix inversion.

To address this problem, we developed a new computational algorithm that converges to the same solution without requiring any matrix inversions. This inverse-free algorithm is conceptually related to the iterative Schulz method for matrix inversion. In the Schulz method, an estimate $P^{(i-1)}$ for the inverse of matrix Q is iteratively refined by¹⁰²

$$P^{(i)} = 2P^{(i-1)} - P^{(i-1)}QP^{(i-1)} \quad (82)$$

where index i is the Schulz iteration. The key difference between our inverse-free algorithm and the Schulz matrix inversion method is that we exploit the particular structures of the dipole interaction tensor and multi-body polarizability matrix contraction to enable us to work with Natom scalars for the non-directional screening and $\text{Natom} \times 3 \times 3$ matrices for the directional screening instead of working with $\text{Natom} \times \text{Natom}$ and $3\text{Natom} \times 3\text{Natom}$ matrices for matrix inversions using GEPP or Schulz method. This allows us to reduce the computational cost from cubic scaling (for GEPP or Schulz method using conventional matrix multiplications) to linear in Natoms as the unit cell becomes sufficiently large.

Substituting $Q_{j+1} = D_j + \Delta_j \tau_j$ from Eq. (64) into (82) with $i=0$ gives the first Schulz iteration as

$$P_{j+1}^{(1)} = 2P_{j+1}^{(0)} - P_{j+1}^{(0)}(D_j + \Delta_j \tau_j)P_{j+1}^{(0)} \quad (83)$$

where

$$\mathbf{P}_{j+1}^{(0)} = \text{inv}(\mathbf{D}_j) \quad (84)$$

is the initial estimate for $\text{inv}(\mathbf{Q}_{j+1})$. The subscript $j+1$ is the screening iteration and should not be confused with the Schulz iteration (superscript i). Substituting Eq. (84) into (83) and simplifying yields

$$\mathbf{P}_{j+1}^{(1)} = \text{inv}(\mathbf{D}_j) - \text{inv}(\mathbf{D}_j)(\Delta_j \tau_j) \text{inv}(\mathbf{D}_j) \quad (85)$$

Note that: (1) τ_j includes non-zero blocks only for interacting atom pairs (i.e., atom pairs in the ‘small’ and/or ‘large’ lists), (2) \mathbf{D}_j and $\text{inv}(\mathbf{D}_j)$ are block-diagonal matrices, (3) the non-zero blocks of $\text{inv}(\mathbf{D}_j)$ are the partially screened atomic polarizabilities $\{\alpha_j^A\}$, and (4) the multi-body polarizability matrix \mathbf{P} is only needed in its contracted form as partially screened atomic polarizabilities (Eqs. (44) and (45)). Using these to simplify Eq. (85) yields

$$\alpha_{j+1}^A = \alpha_j^A - \sum_b \left[\left(\frac{\alpha_{\text{in}}^A}{\alpha_{\text{in}}^A + \alpha_{\text{in}}^B} \right) \left(\alpha_j^A (\Delta \tau_j^{A,b}) \alpha_j^b + \alpha_j^b (\Delta \tau_j^{b,A}) \alpha_j^A \right) \right] \quad (86)$$

Three different types of atomic polarizabilities are computed using this inverse-free algorithm. For non-directional screening to compute $\alpha_A^{\text{non-dir}}(\mathbf{u})$, Eq. (86) becomes

$$\alpha_{j+1}^A(\mathbf{u}) = \alpha_j^A(\mathbf{u}) - \sum_{b \in \text{small_list}} \left[\left(\frac{2\alpha_A^{\text{unscreened}}(\mathbf{u})}{\alpha_A^{\text{unscreened}}(\mathbf{u}) + \alpha_B^{\text{unscreened}}(\mathbf{u})} \right) \left(\alpha_j^A(\mathbf{u}) (\Delta_j \tau_{\text{non-dir},j}^{A,b}(\mathbf{u})) \alpha_j^b(\mathbf{u}) \right) \right] \quad (87)$$

$\tau_{\text{non-dir},j}^{A,b}(\mathbf{u})$ is from Eq. (60), where the Gaussian width $\sigma_{AB,j}(\mathbf{u})$ is updated at the start of screening iteration $j+1$ using $\alpha_j^A(\mathbf{u})$ and $\alpha_j^B(\mathbf{u})$ into Eqs. (18)–(19). The process starts with

$$\alpha_{j=1}^A(\mathbf{u}) = \alpha_A^{\text{unscreened}}(\mathbf{u}) \quad (88)$$

and ends with $\alpha_A^{\text{non-dir}}(\mathbf{u})$ as the value on the left-side of Eq. (87) after the last screening increment finishes.

For directional screening to compute $\alpha_A^{\text{screened}}(\mathbf{u})$, Eq. (86) become

$$\mathbf{M1} = (\Delta_j f_{\text{MBSP}}^{A,b} f_{\text{cutoff}}^{A,b} \phi_j^{A,b}(\mathbf{u})) (\vec{\alpha}_j^A(\mathbf{u}) \bullet (\vec{\eta}^{A,b}(\mathbf{u})) \bullet \vec{\alpha}_j^b(\mathbf{u})) \quad (89)$$

$$\mathbf{M2} = \Delta_j \vec{\alpha}_j^A(\mathbf{u}) \bullet \vec{h}^{A,B}(\mathbf{u}) \bullet \vec{\alpha}_j^B(\mathbf{u}) \quad (90)$$

$$\begin{aligned} \vec{\alpha}_{j+1}^A(\mathbf{u}) = \vec{\alpha}_j^A(\mathbf{u}) - \sum_{b \in \text{small_list}} \left[\left(\frac{\alpha_A^{\text{non-dir}}(\mathbf{u})}{\alpha_A^{\text{non-dir}}(\mathbf{u}) + \alpha_B^{\text{non-dir}}(\mathbf{u})} \right) (\mathbf{M1} + \mathbf{M1}^T) \right] \\ - \sum_{B \in \text{large_list}} \left[\left(\frac{\alpha_A^{\text{non-dir}}(\mathbf{u})}{\alpha_A^{\text{non-dir}}(\mathbf{u}) + \alpha_B^{\text{non-dir}}(\mathbf{u})} \right) (\mathbf{M2} + \mathbf{M2}^T) \right] \end{aligned} \quad (91)$$

where $\mathbf{M1}$ and $\mathbf{M2}$ are square matrices with 3 rows and

$$\phi_j^{A,b}(\mathbf{u}) = - \left(\frac{1}{(d_{Ab})^3} \operatorname{erfc} \left(\frac{d_{Ab}}{\sigma_{AB,j}(\mathbf{u})} \right) + \left(\frac{2}{\sqrt{\pi}} \right) \left(\frac{1}{(d_{Ab})^2 \sigma_{AB,j}(\mathbf{u})} + \frac{2}{3(\sigma_{AB,j}(\mathbf{u}))^3} \right) \exp \left(- \left(\frac{d_{Ab}}{\sigma_{AB,j}(\mathbf{u})} \right)^2 \right) \right) \quad (92)$$

$$\alpha_j^A(\mathbf{u}) = \operatorname{trace}(\vec{\vec{\alpha}}_j^A(\mathbf{u})) / 3 \quad (93)$$

In Eq. (92), the Gaussian width $\sigma_{AB,j}(\mathbf{u})$ is updated at the start of screening iteration $j+1$ using $\alpha_j^A(\mathbf{u})$ and $\alpha_j^B(\mathbf{u})$ from Eq. (93) into Eqs. (18)–(19). The process starts with

$$\vec{\vec{\alpha}}_{j=1}^A(\mathbf{u}) = \begin{bmatrix} \alpha_A^{\text{non-dir}}(\mathbf{u}) & 0 & 0 \\ 0 & \alpha_A^{\text{non-dir}}(\mathbf{u}) & 0 \\ 0 & 0 & \alpha_A^{\text{non-dir}}(\mathbf{u}) \end{bmatrix} \quad (94)$$

and ends with $\alpha_A^{\text{screened}}(\mathbf{u})$ as one-third the trace of the tensor on the left-side of Eq. (91) after the last screening increment finishes.

For directional screening to compute $\vec{\vec{\alpha}}_A^{\text{static}}$, Eq. (86) become

$$\mathbf{M1} = (\Delta_j f_{\text{cutoff}}^{A,b} \phi_j^{A,b}(\mathbf{u})) (\vec{\vec{\alpha}}_j^A(\mathbf{u}) \bullet (\vec{\vec{\eta}}^{A,b}(\mathbf{u})) \bullet \vec{\vec{\alpha}}_j^b(\mathbf{u})) \quad (95)$$

$$\mathbf{M2} = \Delta_j \vec{\vec{\alpha}}_j^A(\mathbf{u}) \bullet \vec{\vec{g}}^{A,B}(\mathbf{u}) \bullet \vec{\vec{\alpha}}_j^B(\mathbf{u}) \quad (96)$$

$$\begin{aligned} \vec{\vec{\alpha}}_{j+1}^A &= \vec{\vec{\alpha}}_j^A - \sum_{b \in \text{small_list}} \left[\left(\frac{\alpha_A^{\text{forcefield}}}{\alpha_A^{\text{forcefield}} + \alpha_B^{\text{forcefield}}} \right) (\mathbf{M1} + \mathbf{M1}^T) \right] \\ &\quad - \sum_{B \in \text{large_list}} \left[\left(\frac{\alpha_A^{\text{forcefield}}}{\alpha_A^{\text{forcefield}} + \alpha_B^{\text{forcefield}}} \right) (\mathbf{M2} + \mathbf{M2}^T) \right] \end{aligned} \quad (97)$$

where $\phi_j^{A,b}(\mathbf{u})$ and $\alpha_j^A(\mathbf{u})$ are defined analogous to Eqs. (92)–(93) except based on $\vec{\vec{\alpha}}_j^A$ from Eq. (97). The process starts using Eq. (94) and ends with $\vec{\vec{\alpha}}_A^{\text{static}}$ as the tensor on the left-side of Eq. (97) after the last screening increment finishes. Summing $\vec{\vec{\alpha}}_A^{\text{static}}$ over all atoms in the unit cell gives the unit cell (or molecular) polarizability tensor. Taking one-third the trace of the atomic or molecular static polarizability tensors yields the corresponding isotropic static polarizabilities.

We now derive the order of the error residual. Since $\mathbf{P}_{j+1}^{(1)}$ is a calculated estimate of $\operatorname{inv}(\mathbf{Q}_{j+1})$, we begin by defining the error residual as

$$\text{residual} = \|\mathbf{I} - \mathbf{Q}_{j+1} \mathbf{P}^{(1)}\| \quad (98)$$

where $\|\cdot\|$ is any desired matrix norm and \mathbf{I} is the identity matrix. Multiplying Eq. (85) by $\mathbf{Q}_{j+1} = \mathbf{D}_j + \Delta_j \tau_j$ gives

$$\mathbf{Q}_{j+1} \mathbf{P}_{j+1}^{(1)} = \mathbf{I} - \Delta_j \tau_j \operatorname{inv}(\mathbf{D}_j) (\Delta_j \tau_j) \operatorname{inv}(\mathbf{D}_j) \quad (99)$$

Multiplying each side of Eq. (99) by -1 , adding \mathbf{I} , and taking the norm gives

$$\|\mathbf{I} - \mathbf{Q}_{j+1} \mathbf{P}^{(1)}\| = \|\Delta_j \tau_j \operatorname{inv}(\mathbf{D}_j) \Delta_j \tau_j \operatorname{inv}(\mathbf{D}_j)\| = (\Delta_j)^2 \|\tau_j \operatorname{inv}(\mathbf{D}_j) \tau_j \operatorname{inv}(\mathbf{D}_j)\| \quad (100)$$

The right-most side follows from the theorem that $\text{norm}(\text{scalar} \times \text{matrix}) = \text{scalar} \times \text{norm}(\text{matrix})$.¹⁰³ For convenience, we used fixed size screening increments: $\Delta_j = \Delta$. Eq. (100) shows that at each screening iteration, the difference between this inverse-free algorithm and direct matrix inversion will be on the order of Δ^2 , where Δ is the screening increment. Since the total number of screening iterations is $1/\Delta$, the total difference between this inverse-free algorithm and direct matrix inversion will be on the order of Δ .

By taking the limit $\Delta \rightarrow 0$, the inverse-free and direct matrix inversion algorithms converge to the identical solution. We used Richardson extrapolation^{98, 104} to evaluate the limit $\Delta \rightarrow 0$. As explained in the prior paragraph, without Richardson extrapolation the overall error using screening increment Δ is of $\text{Order}(\Delta)$. Each Richardson extrapolation step removes a successive power of Δ in the error. After K Richardson extrapolation steps (RES), the remaining error will thus be of $\text{Order}(\Delta^{K+1})$. We extrapolated using screening increments of $1, 2^{-1}, 2^{-2}, \dots, 2^{-K}$. Because the screening increments sum to 1 for a full screening calculation (Eq. (66)), this extrapolation corresponds to extrapolating from results computed with $1, 2, 2^2, \dots, 2^K$ screening points. The final $\Delta=2^{-K}$ undergoes K RES leading to a residual error of $\text{Order}(2^{-K(K+1)})$. Note that 7 RES are approximately twice as expensive as 6 RES and four times as expensive as 5 RES.

Richardson extrapolation was applied to the following atomic polarizabilities:

$$\text{non-directional screening: } \alpha_A^{\text{non-dir}}(\mathbf{u}) = \sum_{\xi=1}^{K+1} c_{(K+1),\xi}^{\text{extrap}} \alpha_A^{\text{non-dir},\xi}(\mathbf{u}) \quad (101)$$

$$\text{frequency-dependent directional screening: } \alpha_A^{\text{screened}}(\mathbf{u}) = \sum_{\xi=1}^{K+1} c_{(K+1),\xi}^{\text{extrap}} \alpha_A^{\text{screened},\xi}(\mathbf{u}) \quad (102)$$

$$\text{static induced directional screening: } \vec{\alpha}_A^{\text{static}} = \sum_{\xi=1}^{K+1} c_{(K+1),\xi}^{\text{extrap}} \vec{\alpha}_A^{\text{static},\xi} \quad (103)$$

As explained above, $\alpha_A^{\text{forcefield},\xi}$, $\alpha_A^{\text{screened},\xi}(\mathbf{u})$, $\vec{\alpha}_A^{\text{static},\xi}$ are the values computed using $2^{\xi-1}$ screening points (i.e., screening increment of $2^{1-\xi}$). The coefficients $c_{(K+1),\xi}^{\text{extrap}}$ for $K=1$ to 7 RES are given in the ESI[†]. Note that

$$\alpha^{\text{low-freq}} = \alpha_A^{\text{screened}}(\mathbf{u} = \text{Nimfreqs}) \quad (104)$$

$$\alpha^{\text{forcefield}} = \alpha_A^{\text{non-dir}}(\mathbf{u} = \text{Nimfreqs}) \quad (105)$$

Figure 12, Figure 13, and Figure 14 illustrate pseudocodes for Richardson extrapolation with inverse-free algorithms used to compute $\{\alpha_A^{\text{non-dir}}(\mathbf{u})\}$, $\{\alpha_A^{\text{screened}}(\mathbf{u})\}$, and $\{\vec{\alpha}_A^{\text{static}}\}$, respectively. The array indices in these figures correspond to what is actually needed in the computer program: (1) the atom number appears as an array index rather than as a subscript, (2) the screening increment does not appear as an array index because temporary results for each screening increment are computed in-place, (3) the imfreq integration point \mathbf{u} appears as an array index only where needed, etc.

Note: This algorithm starts with the unscreened m -scaled results (i.e., $\{\alpha^{\text{unscreened}}(\Lambda)\}$ and $\{\text{wp}(\Lambda)\}$) and small list array as inputs, with $\Lambda = 2^{1/6} \pi^{-1/6} 3^{-1/3}$ and $\gamma = 4/(3\sqrt{\pi})$ already assigned.

Begin non-directional screening

Initialize $\{\alpha^{\text{non-dir}}(\Lambda, u)\} = 0$

Loop over the imfreq points $u = 1, 2, \dots \text{Nimfreqs}$

Calculate $\omega(u) = (\text{Nimfreqs} / u) - 1$

For all atoms, initialize $\alpha^{\text{partial}}(\Lambda) = \alpha^{\text{unscreened}}(\Lambda) / \left(1 + (\omega(u) / \text{wp}(\Lambda))^2\right)$

For each atom image pair (Λ, b) in the small list, calculate and store in the small list array:

$\text{pair_ratio}_{\Lambda b} = \alpha^{\text{partial}}(\Lambda) / (\alpha^{\text{partial}}(\Lambda) + \alpha^{\text{partial}}(B))$

Loop over the Richardson extrapolation steps $s = 0, 1, \dots K$

Calculate $\Delta = 1 / 2^s$

For all atoms, initialize $\alpha^{\text{partial}}(\Lambda) = \alpha^{\text{unscreened}}(\Lambda) / \left(1 + (\omega(u) / \text{wp}(\Lambda))^2\right)$

Loop over the screening increments $j=1, 2, \dots 2^s$

Initialize $\text{temp_vector}(:) = 0$, where $\text{length}(\text{temp_vector}) = \text{Natoms}$

For all atoms, calculate $\sigma(\Lambda) = (\alpha^{\text{partial}}(\Lambda))^{1/3} \Lambda$

Loop over all atom image pairs $\{(\Lambda, b)\}$ in the 'small' list

Read atom numbers for A & B, distance between atom A & image b, r^{cutoff} , and pair_ratio from the small list array

Calculate $\sigma^{\text{pair}} = \sqrt{(\sigma(\Lambda))^2 + (\sigma(B))^2}$

Calculate $x^{\text{pair}} = \text{distance} / \sigma^{\text{pair}}$

If $x^{\text{pair}} > 5$, then cycle

Calculate $\text{temp} = (2r^{\text{cutoff}} \gamma \Delta) \exp\left(-\left(x^{\text{pair}}\right)^2\right) \alpha^{\text{partial}}(\Lambda) \alpha^{\text{partial}}(B) / \left((\sigma^{\text{pair}})^3\right)$

Calculate $\text{temp_vector}(\Lambda) = \text{temp_vector}(\Lambda) + (\text{pair_ratio})(\text{temp})$

Calculate $\text{temp_vector}(B) = \text{temp_vector}(B) + (1 - \text{pair_ratio})(\text{temp})$

End loop

For all atoms: $\alpha^{\text{partial}}(\Lambda) = \alpha^{\text{partial}}(\Lambda) - \text{temp_vector}(\Lambda)$

End loop

For all atoms: $\alpha^{\text{non-dir}}(\Lambda, u) = \alpha^{\text{non-dir}}(\Lambda, u) + c^{\text{extrap}}(K+1, s+1) \alpha^{\text{partial}}(\Lambda)$

End loop

End loop

Apply conduction limit upper bound for all atoms and imfreq points:

$\alpha^{\text{non-dir}}(\Lambda, u) = \text{smooth_min}\left(\alpha^{\text{upperbound}}(\Lambda) / \left(1 + (\omega(u) / \text{wp}(\Lambda))^2\right), \alpha^{\text{non-dir}}(\Lambda, u)\right)$

Assign $\alpha^{\text{force-field}}(\Lambda) = \alpha^{\text{non-dir}}(\Lambda, \text{Nimfreqs})$

End non-directional screening

Figure 12: Pseudo-code for non-directional screening to compute $\{\alpha_A^{\text{non-dir}}(u)\}$ and the force-field polarizabilities $\{\alpha_A^{\text{forcefield}}\}$ using inverse-free algorithm and Richardson extrapolation

Note 1: This algorithm starts with the non-directional screened results (i.e., $\{\alpha^{\text{non-dir}}(A, u)\}$), small list array, and large list array as inputs, with $\Lambda = 2^{1/6} \pi^{-1/6} 3^{-1/3}$ and $\gamma = 4/(3\sqrt{\pi})$ already assigned. Note 2: Because each tensor is symmetric, only six components are unique and required to be stored: $xx, xy=yx, xz=zx, yy, yz=zy, zz$.
Begin imfreq directional screening

Initialize $\{\alpha^{\text{screened}}(A, u)\} = 0$ (See note 2 above.)

Loop over the imfreq points $u = 1, 2, \dots, \text{Nimfreqs}$

For each atom image pair (A, b) in the small list, calculate and store in the small list array:

$$\text{pair_ratio}_{Ab} = \alpha^{\text{non-dir}}(A, u) / (\alpha^{\text{non-dir}}(A, u) + \alpha^{\text{non-dir}}(B, u))$$

For each atom pair (A, B) in the large list, calculate and store in the large list array:

$$\text{pair_ratio}_{AB} = \alpha^{\text{non-dir}}(A, u) / (\alpha^{\text{non-dir}}(A, u) + \alpha^{\text{non-dir}}(B, u))$$

Loop over the Richardson extrapolation steps $s = 0, 1, \dots, K$

Calculate $\Delta = 1/2^s$

For all atoms, initialize $\alpha^{\text{partial}}(A) = \alpha^{\text{non-dir}}(A, u)$,

$$\vec{\alpha}^{\text{partial}}(A) = \text{diag}(\alpha^{\text{partial}}(A), \alpha^{\text{partial}}(A), \alpha^{\text{partial}}(A))$$

Loop over the screening increments $j=1, 2, \dots, 2^s$

For all atoms, initialize $\vec{t}(A) = 0$ (See note 2 above.)

For all atoms, calculate $\sigma(A) = (\alpha^{\text{partial}}(A))^{1/3} \Lambda$

Loop over all atom image pairs $\{(A, b)\}$ in the 'small' list

Read atom numbers for A & B , distance between atom A & image b , f^{cutoff} , f^{MBS} ,
pair_ratio, and $\vec{\eta}^{AB}$ from the small list array

$$\text{Calculate } \sigma^{\text{pair}} = \sqrt{(\sigma(A))^2 + (\sigma(B))^2}$$

$$\text{Calculate } x^{\text{pair}} = \text{distance} / \sigma^{\text{pair}}$$

If $x^{\text{pair}} > 5$, then cycle

Calculate:

$$\text{temp} = (f^{\text{cutoff}} f^{\text{MBS}} \Delta) \left(\frac{\text{erfc}(x^{\text{pair}})}{\text{distance}^3} + (\gamma) \left(\frac{3}{2\sigma^{\text{pair}} \text{distance}^2} + \frac{1}{(\sigma^{\text{pair}})^3} \right) \exp\left(-\left(x^{\text{pair}}\right)^2\right) \right)$$

$$\text{Calculate } \vec{\vec{w}} = \vec{\alpha}^{\text{partial}}(A) \bullet \vec{\eta}^{AB} \bullet \vec{\alpha}^{\text{partial}}(B)$$

$$\text{Calculate } \vec{\vec{w}} = (\vec{\vec{w}} + \text{transpose}(\vec{\vec{w}})) \text{temp}$$

$$\text{Calculate } \vec{t}(A) = \vec{t}(A) + (\text{pair_ratio}) \vec{\vec{w}}$$

$$\text{Calculate } \vec{t}(B) = \vec{t}(B) + (1 - \text{pair_ratio}) \vec{\vec{w}}$$

End loop

Loop over all atom pairs $\{(A, B)\}$ in the 'large' list

Read atom numbers for A & B , pair_ratio, and \vec{h}^{AB} from the large list array

$$\text{Calculate } \vec{\vec{w}} = \vec{\alpha}^{\text{partial}}(A) \bullet \vec{h}^{AB} \bullet \vec{\alpha}^{\text{partial}}(B)$$

$$\text{Calculate } \vec{\vec{w}} = (\vec{\vec{w}} + \text{transpose}(\vec{\vec{w}})) \Delta$$

$$\text{Calculate } \vec{t}(A) = \vec{t}(A) + (\text{pair_ratio}) \vec{\vec{w}}$$

$$\text{Calculate } \vec{t}(B) = \vec{t}(B) + (1 - \text{pair_ratio}) \vec{\vec{w}}$$

End loop

$$\text{For all atoms: } \vec{\alpha}^{\text{partial}}(A) = \vec{\alpha}^{\text{partial}}(A) - \vec{t}(A)$$

$$\text{For all atoms: } \alpha^{\text{partial}}(A) = \text{trace}(\vec{\alpha}^{\text{partial}}(A)) / 3$$

End loop

$$\text{For all atoms: } \alpha^{\text{screened}}(A, u) = \alpha^{\text{screened}}(A, u) + c^{\text{extrap}}(K+1, s+1) \alpha^{\text{partial}}(A)$$

End loop

End loop

$$\text{For all atoms: } \alpha^{\text{low-freq}}(A) = \alpha^{\text{screened}}(A, \text{Nimfreqs})$$

End imfreq directional screening

Figure 13: Pseudocode for imfreq directional screening to compute $\{\alpha_A^{\text{screened}}(u)\}$ and $\{\alpha_A^{\text{low-freq}}\}$ using inverse-free algorithm and Richardson extrapolation

Note 1: This algorithm starts with the non-directional screened results (i.e., $\{\alpha^{\text{force-field}}(\mathbf{A})\}$), small list array, and large list array as inputs, with $\Lambda = 2^{1/6} \pi^{-1/6} 3^{-1/3}$ and $\gamma = 4/(3\sqrt{\pi})$ already assigned. Note 2: Because each tensor is symmetric, only six components are unique and required to be stored: $xx, xy=yx, xz=zx, yy, yz=zy, zz$.

Begin directional static screening

Initialize $\{\vec{\alpha}^{\text{static}}(\mathbf{A})\} = 0$ (See note 2 above.)

For each atom image pair (\mathbf{A}, \mathbf{b}) in the small list, calculate and store in the small list array:

$$\text{pair_ratio}_{\mathbf{Ab}} = \alpha^{\text{force-field}}(\mathbf{A}) / (\alpha^{\text{force-field}}(\mathbf{A}) + \alpha^{\text{force-field}}(\mathbf{B}))$$

For each atom pair (\mathbf{A}, \mathbf{B}) in the large list, calculate and store in the large list array:

$$\text{pair_ratio}_{\mathbf{AB}} = \alpha^{\text{force-field}}(\mathbf{A}) / (\alpha^{\text{force-field}}(\mathbf{A}) + \alpha^{\text{force-field}}(\mathbf{B}))$$

Loop over the Richardson extrapolation steps $s = 0, 1, \dots, K$

Calculate $\Lambda = 1/2^s$

For all atoms, initialize $\alpha^{\text{partial}}(\mathbf{A}) = \alpha^{\text{force-field}}(\mathbf{A})$,

$$\vec{\alpha}^{\text{partial}}(\mathbf{A}) = \text{diag}(\alpha^{\text{partial}}(\mathbf{A}), \alpha^{\text{partial}}(\mathbf{A}), \alpha^{\text{partial}}(\mathbf{A}))$$

Loop over the screening increments $j=1, 2, \dots, 2^s$

For all atoms, initialize $\vec{\mathbf{t}}(\mathbf{A}) = 0$ (See note 2 above.)

For all atoms, calculate $\sigma(\mathbf{A}) = (\alpha^{\text{partial}}(\mathbf{A}))^{1/3} \Lambda$

Loop over all atom image pairs $\{(\mathbf{A}, \mathbf{b})\}$ in the ‘small’ list

Read atom numbers for \mathbf{A} & \mathbf{B} , distance between atom \mathbf{A} & image \mathbf{b} , r^{cutoff} , pair_ratio, and $\vec{\eta}^{\mathbf{AB}}$ from the small list array

$$\text{Calculate } \sigma^{\text{pair}} = \sqrt{(\sigma(\mathbf{A}))^2 + (\sigma(\mathbf{B}))^2}$$

$$\text{Calculate } x^{\text{pair}} = \text{distance} / \sigma^{\text{pair}}$$

If $x^{\text{pair}} > 5$, then cycle

Calculate

$$\text{temp} = (r^{\text{cutoff}} \Delta) \left(\frac{\text{erfc}(x^{\text{pair}})}{\text{distance}^3} + (\gamma) \left(\frac{3}{2\sigma^{\text{pair}} \text{distance}^2} + \frac{1}{(\sigma^{\text{pair}})^3} \right) \exp\left(-\left(x^{\text{pair}}\right)^2\right) \right)$$

$$\text{Calculate } \vec{\mathbf{w}} = \vec{\alpha}^{\text{partial}}(\mathbf{A}) \cdot \vec{\eta}^{\mathbf{AB}} \cdot \vec{\alpha}^{\text{partial}}(\mathbf{B})$$

$$\text{Calculate } \vec{\tilde{\mathbf{w}}} = (\vec{\mathbf{w}} + \text{transpose}(\vec{\mathbf{w}})) \text{temp}$$

$$\text{Calculate } \vec{\mathbf{t}}(\mathbf{A}) = \vec{\mathbf{t}}(\mathbf{A}) + (\text{pair_ratio}) \vec{\tilde{\mathbf{w}}}$$

$$\text{Calculate } \vec{\mathbf{t}}(\mathbf{B}) = \vec{\mathbf{t}}(\mathbf{B}) + (1 - \text{pair_ratio}) \vec{\tilde{\mathbf{w}}}$$

End loop

Loop over all atom pairs $\{(\mathbf{A}, \mathbf{B})\}$ in the ‘large’ list

Read atom numbers for \mathbf{A} & \mathbf{B} , pair_ratio, and $\vec{\mathbf{g}}^{\mathbf{AB}}$ from the large list array

$$\text{Calculate } \vec{\mathbf{w}} = \vec{\alpha}^{\text{partial}}(\mathbf{A}) \cdot \vec{\mathbf{g}}^{\mathbf{AB}} \cdot \vec{\alpha}^{\text{partial}}(\mathbf{B})$$

$$\text{Calculate } \vec{\tilde{\mathbf{w}}} = (\vec{\mathbf{w}} + \text{transpose}(\vec{\mathbf{w}})) \Delta$$

$$\text{Calculate } \vec{\mathbf{t}}(\mathbf{A}) = \vec{\mathbf{t}}(\mathbf{A}) + (\text{pair_ratio}) \vec{\tilde{\mathbf{w}}}$$

$$\text{Calculate } \vec{\mathbf{t}}(\mathbf{B}) = \vec{\mathbf{t}}(\mathbf{B}) + (1 - \text{pair_ratio}) \vec{\tilde{\mathbf{w}}}$$

End loop

$$\text{For all atoms: } \vec{\alpha}^{\text{partial}}(\mathbf{A}) = \vec{\alpha}^{\text{partial}}(\mathbf{A}) - \vec{\mathbf{t}}(\mathbf{A})$$

$$\text{For all atoms: } \alpha^{\text{partial}}(\mathbf{A}) = \text{trace}(\vec{\alpha}^{\text{partial}}(\mathbf{A})) / 3$$

End loop

$$\text{For all atoms: } \vec{\alpha}^{\text{static}}(\mathbf{A}) = \vec{\alpha}^{\text{partial}}(\mathbf{A}) + c^{\text{extrap}}(K+1, s+1) \vec{\alpha}^{\text{partial}}(\mathbf{A}) \text{ (See note 2 above.)}$$

End loop

$$\text{For all atoms: } \alpha^{\text{static}}(\mathbf{A}) = \text{trace}(\vec{\alpha}^{\text{static}}(\mathbf{A})) / 3$$

$$\text{Calculate unit cell polarizability: } \vec{\alpha}^{\text{unit_cell}} = \sum_{\mathbf{A}} \vec{\alpha}^{\text{static}}(\mathbf{A}) \text{ and } \alpha^{\text{unit_cell}} = \sum_{\mathbf{A}} \alpha^{\text{static}}(\mathbf{A})$$

$$\text{Calculate eigenvalues of } \{\vec{\alpha}^{\text{static}}(\mathbf{A})\} \text{ and } \vec{\alpha}^{\text{unit_cell}}$$

End directional static screening

Figure 14: Pseudocode for directional static screening to compute the static polarizability tensors $\{\vec{\alpha}_{\mathbf{A}}^{\text{static}}\}$ using inverse-free algorithm and Richardson extrapolation

Table 8 lists computed results for graphene. After testing, we settled on five RES for the force-field polarizabilities (i.e., non-directional screening), five RES for the low frequency fluctuating polarizabilities used to compute C_6 coefficients, and seven RES for the static polarizabilities. The static polarizability requires more RES than the force-field and low frequency fluctuating polarizabilities, because of the longer range dipole-dipole interactions contributing to the static polarizability. The inverse-free and GEPP algorithms converged to the same results in the limit $\Delta \rightarrow 0$. The inverse-free algorithm is preferable, because it exhibits better computational cost scaling than GEPP.

Table 8: Effect of number of Richardson extrapolations steps (RES) applied to the MCLF screening increments. Graphene was chosen as a test system, because it has strong long-range dipole-dipole coupling.

	(5,5) ^a RES	(6,6) ^a RES	(7,7) ^a RES	(5,7) ^a RES	(7,7) ^a GEPP ^b
α^{static}	20.10	20.14	20.15	20.15	20.15
$\alpha^{\text{low_freq}}$	12.14	12.14	12.14	12.14	12.14
C_6	52.43	52.43	52.43	52.43	52.43
$\alpha^{\text{force-field}}$	7.03	7.03	7.03	7.03	7.03
relative computational cost	17	34	68	20 ^c	cubic

^a The first number in parentheses refers to the number of RES used for non-directional screening and short-range directional screening to compute $\alpha_A^{\text{non-dir}}(\mathbf{u})$ and $\alpha_A^{\text{screened}}(\mathbf{u})$, respectively. The second number in parentheses is the number of RES used for long-range directional screening to compute $\vec{\alpha}_A^{\text{static}}$.

^b Explicit large matrix construction and inversion using Gaussian elimination with partial pivoting. ^c Nominal computational cost for (5,7) algorithm includes 16 frequency points at 5 RES plus the static polarizability at 7 RES, which gives $16(1) + 1(4) = 20$ compared to 17 for the (5,5) algorithm.

Both the required computational time and memory of this inverse-free algorithm are proportional to the number of atoms in the unit cell times the number of separately cataloged pairwise interactions per atom. Case 1: For an isolated molecule much smaller than the dipole interaction cutoff radius, increasing the number of atoms in the molecule also increases the number of pairwise interactions per atom. In this case, the required computational time and memory scale proportional to the number of atoms squared. Case 2: Quadratic scaling of computational time and memory is also observed for periodic materials having small unit cells. As the number of atoms in the unit cell increases, the number of separately cataloged pairwise interactions per atom also increases. Case 3: When the unit cell is large enough to completely enclose a sphere of dipole interaction cutoff radius, the number of separately cataloged pairwise interactions per atom saturates. Making the unit cell even larger does not increase the number of separately cataloged pairwise interactions per atom. In this case, both the required computational time and memory scale linearly with increasing system size.

Figure 15 plots required computational time and random access memory (RAM) to perform MCLF analysis on ice crystals containing different numbers of atoms in the periodic unit cell. These calculations described the same hexagonal ice crystal structure, but with different sized unit cells. MCLF results from these different sized unit cells are numerically equivalent. Electron densities for the unit cells containing 12 to 8748 atoms were taken from the bond order article.⁵⁸ Herein, we also constructed periodic unit cells

containing 20736 to 263424 atoms from the computed DDEC6 atom-in-material properties. As shown in Figure 15, the required computational time and memory for MCLF analysis scaled linearly with increasing number of atoms in the unit cell when the unit cell was large enough to enclose a sphere of dipole interaction cutoff radius.

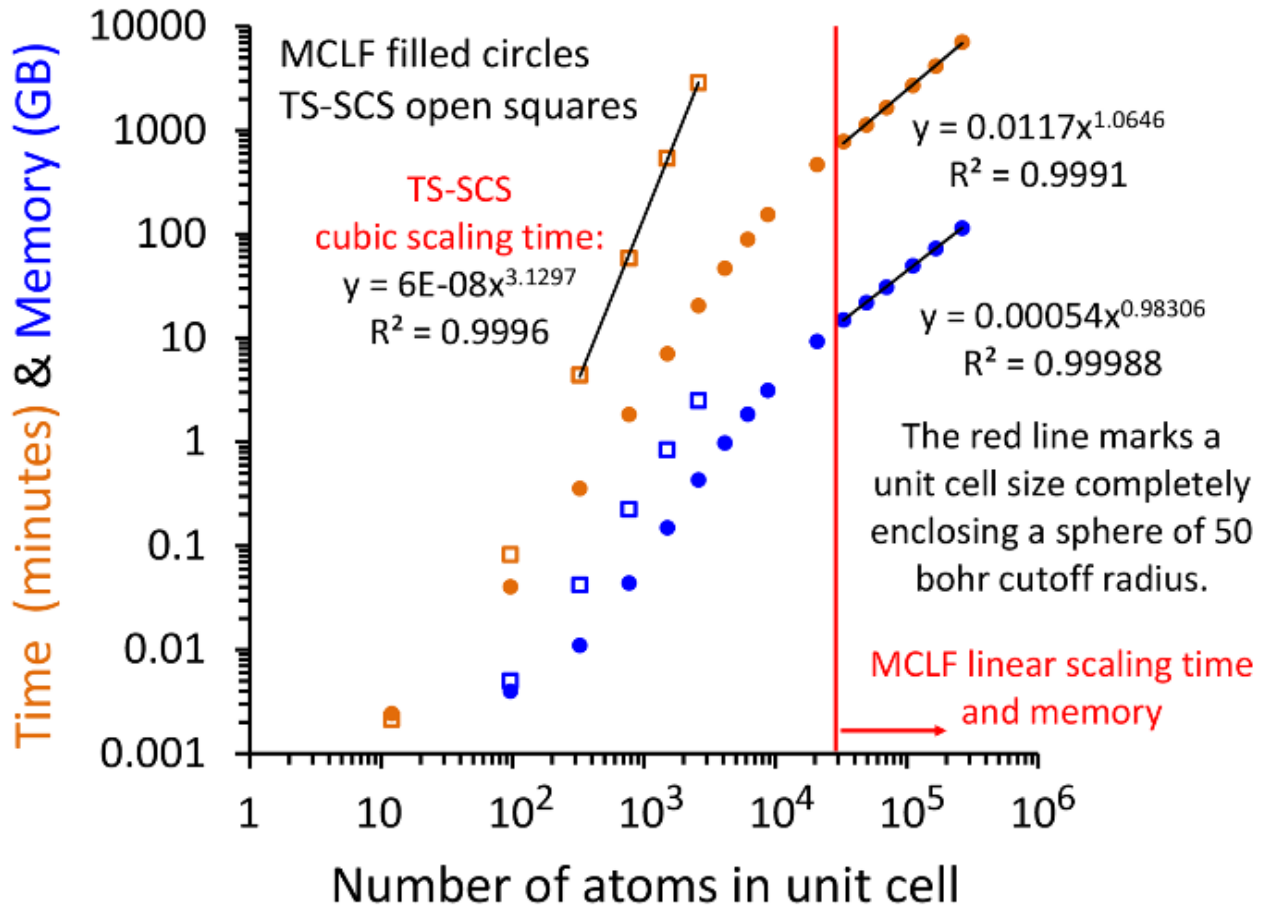


Figure 15: Plot of required computational time and random access memory (RAM) to perform MCLF analysis on ice crystals containing different numbers of atoms in the periodic unit cell: 12, 96, 324, 768, 1500, 2592, 4116, 6144, 8748, 20736, 32928, 49152, 69984, 111132, 165888, and 263424. For 12 atoms, the required RAM was <1 MB. Beyond a certain system size (governed by the dipole interaction cutoff radius), the required computational time and memory scale linearly with increasing system size. TS-SCS results for 12 to 2592 atoms per unit cell are also shown for comparison. Because of direct matrix inversion, the TS-SCS method has nearly cubic scaling computational cost, which makes it infeasible for large unit cells. Each calculation was run in serial mode on a single core in a Xeon E5-2680v3 multi-core processor in the Comet cluster at SDSC.

For comparison, Figure 15 also shows the time and memory (RAM) required to perform TS-SCS analysis on the ice crystals containing 12 to 2592 atoms in the unit cell. To maximize computational efficiency, TS-SCS calculations were performed using small and large list arrays of interacting atom pairs, analogous to those described above for MCLF except collecting the tensor components appropriate for TS-SCS. Same as MCLF, TS-SCS calculations used Rhombert integration with 16 imfreq points. A hard 50 bohr dipole interaction cutoff was used for TS-SCS compared to the smooth cutoff (Eq. (63)) used for MCLF. Because TS-SCS analysis used explicit large matrix construction and inversion,³⁸ TS-SCS analysis required more computational time and memory than MCLF analysis. Because of direct matrix

inversion using GEPP, the TS-SCS method had nearly cubic scaling computational cost with increasing number of atoms in the unit cell. We attempted TS-SCS calculations on ice crystals containing ≥ 4116 atoms per unit cell, but these did not complete within one week. It is currently unclear whether it is future possible to create an iterative-based TS-SCS algorithm with linear scaling computational cost.

6. Results and Discussion

Unless otherwise labeled, the following results are calculated using DDEC6 partitioning. The DDEC6 results were calculated using the Chargemol program.^{52, 57-59}

6.1 Diatomic molecules

We first tested our model's sensitivity to the choice of exchange-correlation functional and basis set used to compute the system's electron and spin density distributions. A set of 57 diatomic molecules with elements across the periodic table were chosen as the test set. Three sets of electron density distributions of the test set were generated using Gaussian09 with CCSD/def2QZVPPDD, Gaussian09 with B3LYP/def2QZVPPDD, and VASP with PBE/planewave. The Chargemol program was then used to DDEC6 partition the electron density followed by MCLF analysis to obtain the static polarizabilities and C_6 coefficients. Figure 16 shows polarizability and C_6 computed using CCSD/def2QZVPPDD densities versus polarizability and C_6 computed using the PBE/planewave and B3LYP/def2QZVPPDD densities. This figure shows our model did not show strong dependence on the choice of exchange-correlation functional or basis set. Hence, MCLF gives similar results using electron density distributions from different proficient quantum chemistry levels of theory.

For the same 57 diatomic molecules, the isotropic static polarizability and the three eigenvalues of the static polarizability tensor from TS-SCS and MCLF were compared to the reference data. The reference is CCSD calculations with def2QZVPPDD basis set. System-specific polarizabilities for CCSD reference, TS-SCS, and MCLF are listed in the ESI†. Table 9 summarizes the error statistics. The TS-SCS method gave large errors independent of the charge partitioning method used (Hirshfeld or DDEC6). On average, MCLF was four times more accurate than TS-SCS for these materials.

Figure 17 is the absolute % error of isotropic polarizability from TS-SCS vs. NAC of the atoms in the molecules. We only use the positive NAC to represent the whole molecule because for neutral diatomic molecules, the positive and negative charged atom have the same absolute value of the charge. From the plot, we can see TS-SCS gives much larger errors for highly charged molecules than neutral and low-charge molecules. This confirms that TS model does not work for charged atoms and MCLF fixed the problem with highly charged systems.

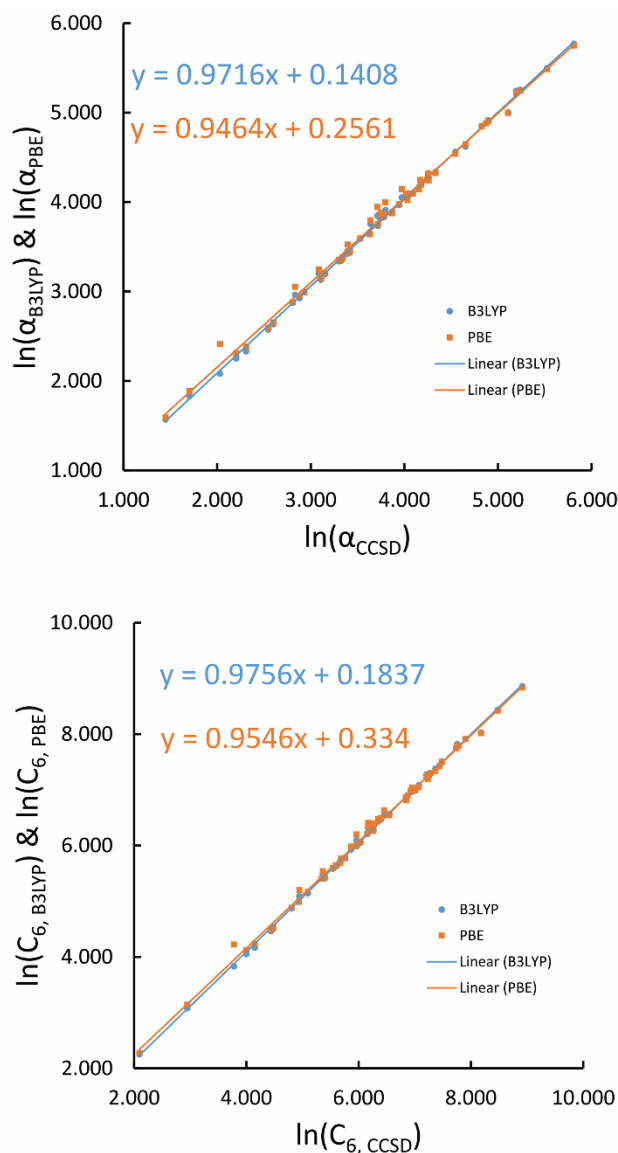


Figure 16: Comparison of MCLF α and C_6 of 57 diatomic molecules computed using electron and spin density distributions from CCSD, PBE and B3LYP calculations with large basis sets.

Table 9: Comparison of the % error in the static polarizability of 57 diatomic molecules. The isotropic static polarizability equals one-third of the trace of the static polarizability tensor.

	TS-SCS(H)		TS-SCS(DDEC6)		MCLF	
	isotropic	eigenvalues	isotropic	eigenvalues	isotropic	eigenvalues
MRE	34.43%	36.08%	27.83%	29.36%	-7.78%	-7.96%
MARE	41.06%	44.92%	36.56%	40.23%	10.44%	13.67%
Range	-21 – 437%	-35 – 484%	-24 – 440%	-35 – 491%	-34 – 17%	-40 – 52%

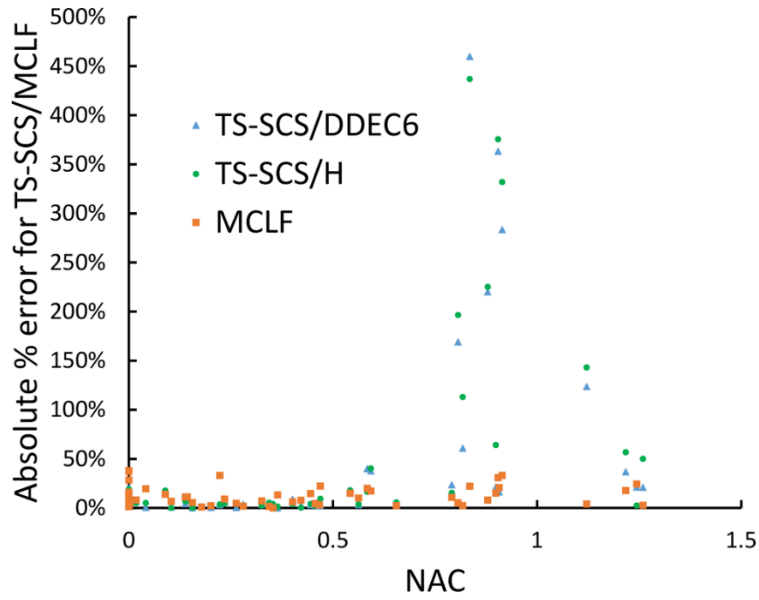


Figure 17: The absolute % error of isotropic static polarizabilities from TS-SCS/H, TS-SCS/DDEC6 and MCLF versus DDEC6 NAC magnitude of 57 diatomic molecules.

6.2 Dense periodic solids

Static polarizabilities were computed for 28 dense periodic solids including electric insulators, semi-conductors, and conductors. The geometries are from Inorganic Crystal Structure Database (ICSD).¹⁰⁵ We generated electron densities in VASP¹⁰⁶ using the PBE¹⁰⁷ functional. See Gabaldon-Limas and Manz⁵⁹ for a description of the VASP computational settings used.

Since the polarizability should not exceed the conduction limit, the reference static polarizability was set equal to the smaller value of the Clausius-Mosotti relation (eq. (106)) and conduction limit (eq. (107)):

$$\alpha = \frac{\text{volume}}{\text{atom}} * \frac{3}{4\pi} * \frac{\kappa - 1}{\kappa + 2} \quad (106)$$

$$\alpha = \frac{\text{volume}}{\text{atom}} * \frac{1}{2\pi} \quad (107)$$

where volume is obtained from the Inorganic Crystal Structure Database (ICSD¹⁰⁸) crystal structure and κ is the experimental static dielectric constant. Results from different partitioning methods and screening methods were compared this reference data. ICSD codes and computed results for individual materials are listed in the ESI†.

Table 10 summarizes the error statistics. Comparing the MRE and MARE of the screened and unscreened polarizabilities with the same partitioning method (e.g., TS/H vs. TS-SCS/H), screening increases the accuracy for each partitioning method. Because all of the unscreened methods gave >100% error for some materials, screening is an essential step in polarizability calculations. Comparing TS-SCS with different partitioning methods: IH is more accurate than H, and DDEC6 is more accurate than IH. The tendency of TS-SCS/H and TS-SCS/IH to overestimate polarizabilities for solids was previously reported by Bucko et al.⁴⁸ with improved results reported using the fractionally ionic approach of Gould et al.⁴⁵ (Those studies included some of the same materials as here.^{45, 48}) Using DDEC6 partitioning, MCLF gives more accurate results than TS-SCS with MARE of 12% and 24%, respectively.

Table 10: Comparison of the % error in the computed static polarizabilities of 28 solids using different methods. H indicates Hirshfeld partitioning. IH indicates iterative Hirshfeld partitioning.

	<i>unscreened methods</i>			
	TS (H)	TS (IH)	TS (DDEC6)	unscreened MCLF
MRE	297%	141%	86%	36%
MARE	297%	152%	93%	55%
Range	11–714%	-53–537%	-28–379%	-36–180%
	<i>screened methods</i>			
	TS-SCS (H)	TS-SCS (IH)	TS-SCS (DDEC6)	MCLF
MRE	46%	16%	15%	-9%
MARE	47%	30%	24%	12%
Range	-10–76%	-53–50%	-31–50%	-37–14%

6.3 Polarizabilities of Small Molecules

A set of 22 molecules were selected from Thole²⁷ and Applequist et al.²⁶ that have experimentally measured isotropic static polarizability. Six of these also have experimentally measured polarizability tensor eigenvalues.^{26,27} The geometries are from geometry optimization we performed in Gaussian09 with B3LYP functional and def2QZVPPDD basis set. Table 11 and Table 12 show that both TS-SCS and MCLF performed well for this test set.

Table 11: Comparison of the isotropic static polarizability in atomic units of 22 molecules

	reference	TS-SCS (DDEC6)	MCLF		reference	TS-SCS (DDEC6)	MCLF
C ₂ H ₂	22.472	23.540	27.855	CH ₃ CN	30.233	33.818	34.111
C ₂ H ₄	28.694	28.886	29.127	(CH ₃) ₂ CO	43.122	47.864	43.785
H ₂ O	9.785	9.168	35.429	CH ₃ OCH ₃	35.361	38.102	36.270
C ₆ H ₆	69.868	70.437	75.646	CH ₂ ClCN	41.165	46.973	19.532
CF ₄	19.705	21.928	23.071	CH ₂ OCH ₂	29.900	32.003	53.566
CFCl ₃	63.907	57.929	61.873	C ₂ H ₅ OH	34.282	37.546	9.845
NH ₃	16.129	13.862	57.983	H ₂ CO	16.533	17.979	18.624
CO ₂	19.644	17.884	49.071	HCONH ₂	27.938	28.328	30.208
CS ₂	59.385	51.175	30.739	CH ₂ Br ₂	60.735	57.041	14.368
C ₃ H ₈	42.829	47.435	42.233	SF ₆	44.134	35.627	37.522
C ₂ H ₆	30.233	32.626	29.131	SO ₂	26.993	25.164	26.650
MRE						1.04%	2.92%
MARE						8.74%	7.49%

Table 12: Comparison of the static polarizability tensor eigenvalues in atomic units of 6 molecules. For each molecule, the three eigenvalues were sorted smallest to largest.

	reference			TS-SCS(DDEC6)			MCLF		
	eigen 1	eigen 2	eigen 3	eigen 1	eigen 2	eigen 3	eigen 1	eigen 2	eigen 3
C ₂ H ₅ OH	30.368	33.607	38.870	30.219	34.512	47.909	30.702	33.116	42.469
CF ₄	19.705	19.705	19.705	21.928	21.928	21.928	23.071	23.071	23.071
C ₂ H ₆	27.668	27.668	35.361	28.210	28.210	41.457	27.495	27.496	32.403
CH ₃ CN	25.981	25.981	38.735	23.995	23.996	53.463	22.735	22.735	56.862
(CH ₃) ₂ CO	31.380	48.959	49.027	35.472	49.630	58.489	33.686	48.247	49.422
CH ₃ OCH ₃	29.625	33.337	43.054	31.823	31.989	50.495	31.998	32.168	44.643
MRE					8.75%			4.85%	
MARE					10.95%			9.40%	

6.4 C₆ Coefficients of Atom/Molecule Pairs

This test involves C₆ coefficients for pairs of atoms and molecules studied by Tkatchenko and Scheffler.¹⁰ The geometries are from geometry optimization we performed in Gaussian 09 with B3LYP functional and def2QZVPPDD basis set.

Figure 18 compares the TS-SCS(DDEC6) and MCLF C₆ coefficients to the reference values derived from the dipole oscillator strength distribution (DOSD) data of Meath and co-workers^{65, 66} as tabulated in the supporting information of Bucko et al.⁴⁶ As shown in Figure 18, TS-SCS predicts too large C₆ values for the larger molecules, while MCLF does not have this problem. For these 49 atoms/molecules, MCLF gave 1.15% MRE and 5.79% MARE while TS-SCS(DDEC6) gave 8.09% MRE and 10.29% MARE.

The same reference data source was also used for the 1225 pairs formed from these 49 atoms/molecules. Figure 19 plots MCLF versus reference C₆ values for these pairs. These MCLF C_{6, AB} values were computed from the C_{6, A} values using the mixing rule in Eq. (81). MCLF yielded highly accurate results with 0.80% MRE and 4.45% MARE. Results for individual materials in this data set are listed in the ESI†.

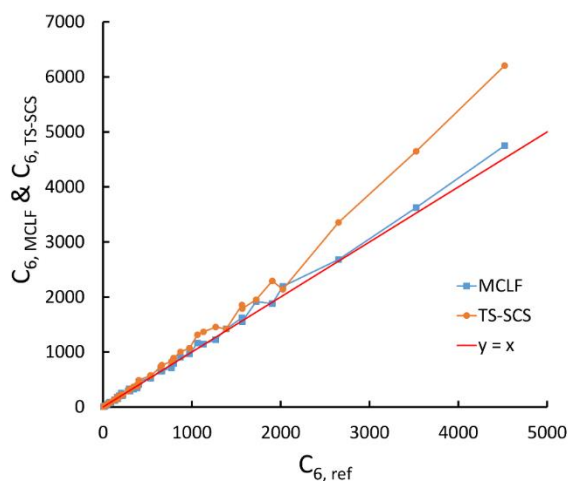


Figure 18: TS-SCS(DDEC6) and MCLF predicted C₆ in atomic units of 49 atoms/molecules compared to experimentally-derived reference C₆ values. TS-SCS(DDEC6) predicts too large C₆ values for the larger molecules.

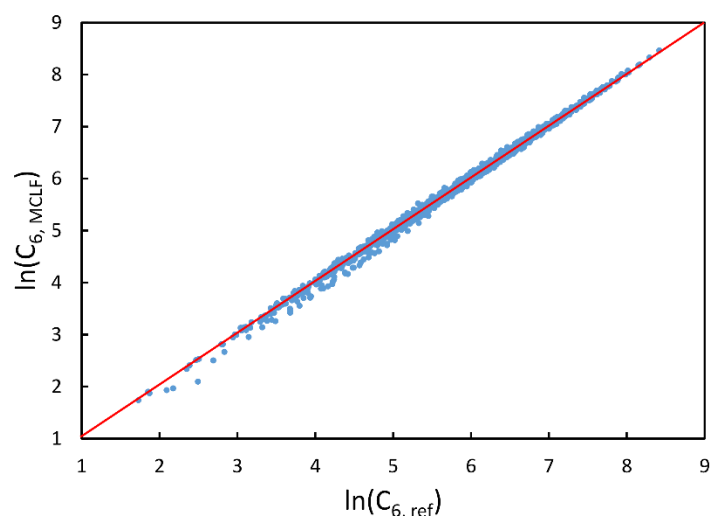


Figure 19: MCLF predicted C_6 coefficients in atomic units for 1225 pairs formed from 49 atoms/molecules compared to the experimentally-derived reference C_6 coefficients.

6.5 Polyacenes and fullerenes

Polycyclic aromatic compounds, such as polyacenes, and fullerenes, have strong directional alignment of dipoles in the ring planes. A set of 12 polyacenes and a set of 6 fullerenes were selected as test sets. We generated electron densities in VASP¹⁰⁶ using the PBE¹⁰⁷ functional. See Gabaldon-Limas and Manz⁵⁹ for a description of the VASP computational settings used.

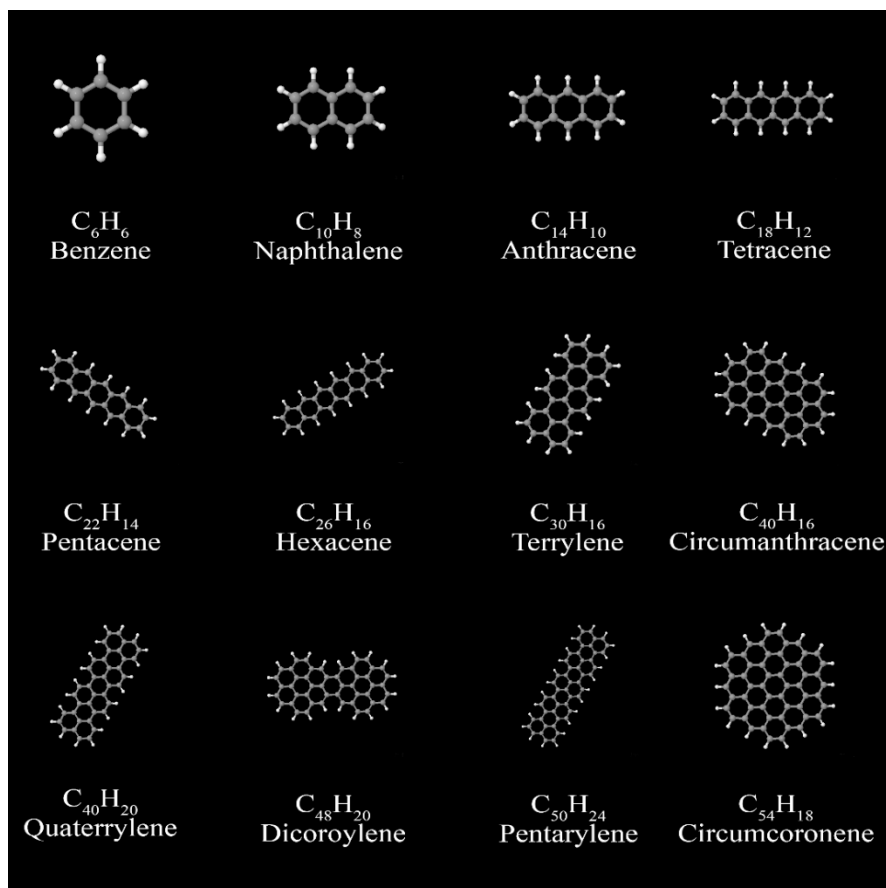


Figure 20: Structures of 12 polyacenes

The polyacene geometries are from our VASP geometry optimization using the PBE functional. Polyacene reference static polarizabilities and C_6 coefficients are from Marques et al.⁶⁸ The polyacene structures are shown in Figure 20. Computed results are summarized in Table 13. The MRE and MARE show MCLF was significantly more accurate than TS-SCS for describing both the static polarizabilities and the C_6 coefficients of these materials.

Table 13: Comparison of the α and C_6 in atomic units for 12 polyacenes

	static polarizability, α			C_6 dispersion coefficient		
	reference	TS-SCS (DDEC6)	MCLF	reference	TS-SCS (DDEC6)	MCLF
C_6H_6	70.5	71.864	79.100	1730	2018.50	1999.34
$C_{10}H_8$	123	122.500	133.052	4790	5641.58	5298.26
$C_{14}H_{10}$	189	181.475	197.738	9920	11866.54	10477.15
$C_{18}H_{12}$	264	247.574	271.016	17500	21191.56	17595.20
$C_{22}H_{14}$	353	319.109	350.487	28100	33936.60	26638.47
$C_{26}H_{16}$	454	395.191	434.779	42100	50407.20	37626.79
$C_{30}H_{16}$	484	402.755	441.127	47800	55318.93	46503.57
$C_{40}H_{16}$	612	523.270	590.151	82500	95014.04	81069.67
$C_{40}H_{20}$	799	570.497	626.426	97000	105993.06	83472.16
$C_{48}H_{20}$	770	665.562	745.500	122000	147407.13	118431.09
$C_{50}H_{24}$	1196	748.553	822.271	168000	175992.78	131300.24
$C_{54}H_{18}$	840	707.953	806.519	150000	173114.17	147130.26
MRE		-13.15%	-4.14%		16.40%	-2.38%
MARE		13.47%	8.75%		16.40%	7.77%

Table 14 summarizes calculation results for fullerenes. The fullerene geometries are from Saidi and Norman.¹⁰⁹ Tao et al. studied this set of fullerenes in 2016 and obtained excellent results using a hollow sphere model with modified single frequency approximation.¹¹⁰ The reference static polarizabilities and C_6 coefficients are from Kauczor et al.'s TD-DFT calculations.¹¹¹ For this test set, TS-SCS systematically underestimates the polarizabilities by 18.67% and C_6 coefficients by 10.44%.

In these two test sets, MCLF has better overall performance than TS-SCS with all four MCLF MAREs under 10% compared to all four TS-SCS MAREs over 10%. In contrast to TS-SCS, MCLF uses (i) an iterative update of the Gaussian dipole width and (ii) a multi-body screening function (eqn (70)) to describe decay of the dipole directional order. These allow MCLF to describe dipole directional alignment effects more precisely than TS-SCS.

Table 14: Comparison of the α and C_6 in atomic units for 6 fullerenes

	static polarizability, α			C_6 dispersion coefficient		
	reference	TS-SCS (DDEC6)	MCLF	reference	TS-SCS (DDEC6)	MCLF
C ₆₀	536.6	446.94	512.624	100100	88860.41	106807.97
C ₇₀	659.1	534.37	616.607	141600	125502.37	150149.97
C ₇₈	748.3	605.63	701.272	178200	159842.09	190785.18
C ₈₀	798.8	626.63	727.065	192500	170229.03	201556.87
C ₈₂	779.7	642.93	746.097	196800	179253.82	213111.45
C ₈₄	806.1	659.27	765.368	207700	188430.29	224796.59
MRE		-18.67%	-5.92%		-10.44%	6.84%
MARE		18.67%	5.92%		10.44%	6.84%

6.6 Large biomolecule

Non-reactive molecular mechanics force fields are model potentials containing several different parameter types: dispersion-repulsion parameters (e.g., Lennard-Jones, Buckingham, or other forms), point charges or model atomic charge distributions (e.g., Gaussian, Slater), flexibility parameters, and (optionally) polarizabilities.^{112, 113} These molecular mechanics force fields enable classical atomistic simulations, such as molecular dynamics or Monte Carlo, to be performed over larger distance and time scales than would be practical using quantum chemistry methods such as density functional theory (DFT).¹¹³⁻¹¹⁵ These atomistic simulations are useful to estimate thermodynamic ensemble properties (e.g., density, vapor pressure, adsorption isotherms, etc.), transport properties (e.g., diffusion coefficients, viscosity, etc.), and structures (e.g., protein folding and other conformational changes).^{113, 114, 116-119}

An approach often used is to classify atoms into types, where similar atoms share the same atom type and same force-field parameters.^{120, 121} We will refer to these as Typed Force Fields (TFF). TFFs have been successively improved over the past few decades by refining their atom type definitions and parameter values to make them more accurate and robust.¹²²⁻¹²⁴ Today, several TFFs perform reasonably well on various organic molecules and some small inorganic molecules.^{117, 118, 125-128}

There are still areas for further improving force fields, especially for systems containing high chemical bonding diversity and charged ions. Metal-containing systems are especially prone to high chemical bonding diversity. For example, metal-organic frameworks (MOFs) can contain dozens of different metal elements in a plethora of different bonding motifs.¹²⁹⁻¹³¹ While some efforts have been made to define new atom types for MOFs,¹³² the high chemical diversity makes it difficult to completely parameterize force fields for all MOFs using atom types. Quantum-mechanically derived force-fields (QMDFFs) are ideal for modeling these systems, because QMDFFs do not require pre-defined atom types.^{133, 134} Machine learning is a recent approach in which force-field parameterization is trained to a machine learning model using QM-derived parameters.^{4, 135, 136} Advantages of the machine learning approach include highly automated, computationally fast, and the ability to handle large chemical diversity without having to manually define atom types.^{4, 135, 136}

Non-polarizable force fields often model charged ions with reduced effective charges, but this places artificial constraints on the simulation.¹³⁷ To make the parameters more transferable between different chemical systems and compositions, the reduced effective charge model should be replaced with a polarizable force-field.^{14, 17, 138-140} Kiss and Baranyai concluded for water that “It is impossible to describe the vapor-liquid coexistence properties consistently with nonpolarizable models, even if their

critical temperature is correct.”¹³ Hence, an automated method like MCLF to assign atom-in-material polarizabilities is extremely important for modeling materials containing ions.

Useful insights can be gained by comparing TFF parameters for an atom type to QM-derived ones. Where the TFF parameters and the QM-derived ones are in good agreement, this validates the parameterization. Conversely, where the TFF parameters and the QM-derived ones differ substantially, this indicates areas for further study to potentially refine the atom type definitions and/or parameter values. A multimodal distribution or wide range of QM-derived parameter values suggests when to divide atoms into multiple atom types. A narrow distribution of QM-derived parameter values that is substantially offset from the TFF parameter value can indicate a need to update the TFF parameter value. In such a way, these comparisons can produce force-field improvements.

In this section, we compare MCLF C_6 atom-in-material dispersion coefficients and polarizabilities to OPLS and AMOEBA force-field parameters for the Human Immunodeficiency Virus reverse-transcriptase (HIV-RT) enzyme complexed with an inhibitor molecule. HIV is a retrovirus with RNA genome.¹⁴¹ Retroviruses replicate in a host cell by using a reverse-transcriptase enzyme to transcribe the virus’s RNA genome into the host cell’s DNA that is subsequently replicated by the host.¹⁴¹ Therefore, inhibition of the reverse-transcriptase enzyme is a potential way to slow virus replication, which is extremely important for controlling disease caused by the virus.¹⁴¹

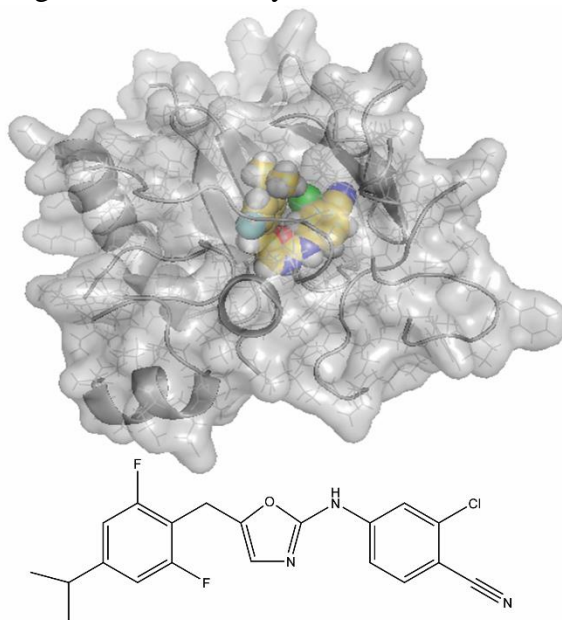


Figure 21: Structure of the inhibitor molecule and its complex with HIV reverse-transcriptase. In the complex, atoms of the inhibitor molecule are displayed as colored balls: yellow (C), white (H), red (O), blue (N), cyan (F), and Cl (green).

Bollini et al. and Cole et al. previously studied several oxazole derivatives as HIV-RT inhibitors.¹⁴²⁻¹⁴⁴ As an example of the MCLF method applied to macromolecules, we study one of these HIV-RT inhibitors, $C_{20}H_{16}ClF_2N_3O$ (CAS # 1422256-80-1), shown in Figure 21 together with a significant portion (2768 atoms) of its complex with wild-type HIV-RT. We computed the electron density distribution for this complex in ONETEP^{145, 146} using the PBE¹⁰⁷ exchange-correlation functional. Preparation of the input structure is described elsewhere.^{142, 143} It was constructed from the 1S9E PDB file¹⁴⁷ using the MCPRO¹⁴⁸ and BOMB¹⁴⁹ software. The 178 amino acids closest to the ligand were retained. The complex was solvated in a 25 Å water cap and equilibrated at room temperature for 40

million Monte Carlo steps using the MCPRO software. Water molecules were stripped from the final configuration, and the resulting structure (Figure 21) was used as input for the ONETEP calculations.

Interactions between electrons and nuclei were described by Opium norm-conserving pseudopotentials.¹⁵⁰ NGWFs were initialized as orbitals obtained from solving the Kohn-Sham equation for isolated atoms.¹⁵¹ The NGWFs were expanded as a psinc basis set¹⁵² with an equivalent plane-wave cut-off energy of approximately 1000 eV and the electron density was stored on a Cartesian grid of spacing 0.23 Bohr. The localization radii of the NGWFs were 10.0 Bohr. Calculations were performed using an implicit solvent model with a dielectric constant of 78.54 to mimic the water environment.¹⁵³ Electron density partitioning was performed using the DDEC6 method implemented in the Chargemol program.^{52, 57-59}

Table 15 compares the computed C_6 coefficients in the HIV-RT complex for a number of frequently-occurring OPLS atom types,¹¹⁷ including both backbone and side chain atoms. The C_6 coefficients computed using MCLF generally show a much smaller range than the corresponding TS-SCS(DDEC6) data. Also, the OPLS force field C_6 coefficients are usually closer to the MCLF results than to the TS-SCS results. While one should not draw too many conclusions from this, the OPLS parameters have been carefully fit over a number of decades to accurately reproduce experimental observables, such as organic liquid properties¹¹⁷ and protein NMR measurements¹⁵⁴. It is also noteworthy that the OPLS C_6 coefficient is slightly higher than the MCLF result. This is expected since the C_6 term in the OPLS force-field must effectively compensate for higher-order dispersion (C_8 , C_{10} , ...) terms that are not explicitly included in the OPLS force-field.¹⁵⁵ The high value of C_6 on the backbone carbonyl carbon (C (bb)) has been noted previously,³ and should be revisited in future force fields.

Table 15. Comparison of QM-derived and TFF C_6 coefficients in atomic units for the HIV reverse transcriptase complex. Two QM-derived methods (MCLF and TS-SCS) are compared to the OPLS TFF. The mean unsigned deviation (MUD) quantifies the QM-derived C_6 coefficient variation compared to the mean value for that atom type. bb signifies backbone atoms.

Atom Type	MCLF		TS-SCS(DDEC6)		OPLS
	Range	Mean (MUD)	Range	Mean (MUD)	
N (bb)	31.7-47.2	39.0 (1.9)	33.7-71.9	48.2 (5.2)	58.2
H (bb)	0.7-1.2	0.9 (0.1)	0.2-1.4	0.6 (0.2)	0.0
C (bb)	20.1-28.9	24.4 (1.0)	19.3-54.8	34.7 (4.1)	84.8
O (bb)	25.3-39.9	31.8 (2.8)	18.4-31.5	24.1 (1.8)	41.0
CT (bb)	27.8-32.7	30.3 (0.9)	50.7-81.2	64.8 (5.0)	35.2
CT (CH3)	30.2-34.7	32.2 (0.8)	45.0-76.0	55.1 (4.2)	35.2
CT (CH)	24.7-37.0	30.5 (2.2)	35.7-98.8	53.8 (8.0)	35.2
HC	1.1-2.6	1.6 (0.2)	0.4-2.1	0.7 (0.2)	2.1
CA	32.6-43.1	36.9 (1.9)	36.0-65.2	45.7 (4.1)	40.7
HA	1.1-2.2	1.6 (0.2)	0.6-2.5	1.3 (0.3)	1.8

Table 16 compares the three kinds of MCLF polarizabilities (i.e., force-field, static, and low freq) to parameters used in the AMOEBA^{14, 139} polarizable force field. For the H atoms, the AMOEBA polarizabilities are larger than the MCLF polarizabilities, though both are small compared to the

polarizabilities of C, N, and O atoms. For the N (bb) atom, the AMOEBA polarizability is similar to the MCLF force-field polarizability. For the O (bb) atom, the AMOEBA polarizability (5.6) is smaller than the MCLF polarizability (6.9). Since the polarizability of an isolated oxygen atom is ~ 5.2 ,⁵⁶ these polarizabilities are in line with O (bb) being slightly negatively charged and more diffuse than an isolated oxygen atom. For the C atoms, the AMOEBA polarizabilities are substantially larger than the MCLF force-field polarizabilities and close to the MCLF static polarizabilities; this suggests the AMOEBA C atom polarizabilities include some directional screening effects. The AMOEBA polarizabilities use Thole²⁷ model atomic charge distributions¹⁴ $c_1 \exp(-c_2(r_A)^3)$ while the MCLF polarizabilities use Gaussian model charge distributions, so the optimized atomic polarizabilities can be different in these two methods and still approximately reproduce the molecular polarizability. Importantly, computed results presented earlier in this article show the MCLF force-field polarizabilities input into the dipole interaction tensor and solved yield good accuracy molecular static polarizabilities. Therefore, the MCLF force-field polarizabilities are appropriate for use in polarizable force-fields.

Table 16. Comparison of MCLF QM-derived and AMOEBA TFF polarizabilities in atomic units for the HIV reverse transcriptase complex. All three MCLF polarizabilities are listed: force-field, static, and low freq. The mean unsigned deviation (MUD) quantifies the MCLF polarizability variation compared to the mean value for that atom type. bb signifies backbone atoms.

Atom Type	MCLF (force-field)		MCLF (static)		MCLF (low freq)		AMOEBA
	Range	Mean (MUD)	Range	Mean (MUD)	Range	Mean (MUD)	
N (bb)	6.4-7.8	7.4 (0.2)	8.8-15.0	11.4 (0.9)	8.1-11.1	9.6 (0.4)	7.2
H (bb)	1.4-2.0	1.6 (0.1)	1.5-2.0	1.7 (0.1)	1.5-2.0	1.7 (0.1)	3.3
C (bb)	5.6-6.4	6.1 (0.1)	7.3-11.5	9.0 (0.5)	7.0-8.6	7.8 (0.2)	9.0
O (bb)	6.2-7.7	6.9 (0.3)	6.3-11.4	8.4 (0.9)	6.5-9.4	7.8 (0.5)	5.6
CT (bb)	6.7-7.3	7.0 (0.1)	8.8-11.3	9.9 (0.4)	8.1-9.0	8.5 (0.2)	9.0
CT (CH ₃)	7.0-8.0	7.6 (0.2)	8.3-10.1	9.0 (0.3)	8.0-8.9	8.4 (0.2)	9.0
CT (CH)	6.6-8.1	7.4 (0.3)	7.8-12.8	9.5 (0.7)	7.5-9.7	8.5 (0.4)	9.0
HC	1.6-2.9	2.1 (0.1)	1.7-3.1	2.2 (0.1)	1.6-3.0	2.1 (0.1)	3.3
CA	7.2-8.3	7.7 (0.2)	9.6-13.1	10.9 (0.6)	8.9-10.5	9.6 (0.3)	11.8
HA	1.8-2.4	2.0 (0.1)	1.9-2.9	2.3 (0.2)	1.9-2.6	2.2 (0.1)	4.7

OPLS atom types¹¹⁷ were assigned using MCPRO software¹⁴⁸ and have the following descriptions. Sidechain atom types included: CT(CH₃): alkane carbon bonded to three hydrogen atoms; CT(CH): alkane carbon atom bonded to one hydrogen atom; HC: alkane hydrogen atom; CA: carbon atom in an aromatic ring; HA: hydrogen bonded to an aromatic ring. Backbone atoms were classified according to: N: amide nitrogen; H: hydrogen bonded to amide nitrogen; C: amide carbon; O: amide oxygen; CT: alpha carbon atom. Only the most common atom types were analyzed to ensure adequate statistics. The substantial ranges of QM-derived parameter values in Table 15 and Table 16 suggest there is room to further improve the atom type definitions. For example, by defining atom types according to the set of directly bonded

neighbors (e.g., N-C2H as an atom type denoting N directly bonded to two carbon atoms, one hydrogen atom, and no others) and (optionally) second bonded neighbors, which was previously shown^{4, 59} to give highly consistent DDEC6 NACs. We recommend the atom type definitions be explored further in future work.

7. Conclusions

In summary, we introduced a new method (MCLF) to compute polarizabilities and dispersion force coefficients. This method can be applied to large and complex materials for which ab initio methods such as time-dependent DFT or CCSD perturbation response theory are too computationally expensive. Like TS-SCS, this method: (i) only requires the electron and spin density distributions as inputs, (ii) is capable of computing polarizability tensors and C_6 coefficients for atoms-in-materials as well as for the whole molecule or unit cell, and (iii) works for materials containing 0, 1, 2, or 3 periodic boundary conditions. We show that using DDEC6 partitioning increases accuracy compared to Hirshfeld and IH partitioning. The MCLF method achieves a long list of important improvements compared to existing methods:

- 1) MCLF has new polarizability and C_6 scaling relations for isolated atoms to set the reference values. These cover the charged atom states using a fundamentally different approach than the Fractional Ionic (FI) and TS-SCS methods. Unlike FI, this new approach does not require quantum mechanically computed reference polarizabilities and C_6 values for charged atoms; this is a huge advantage, because some isolated charged atoms are unstable. Unlike the TS method, this new approach describes changes in the polarizability-to-volume ratio with atomic charge state.
- 2) MCLF uses a new polarizability partition that avoids negative polarizabilities for highly charged atoms (e.g., ZrO molecule) by partitioning the mixed pair contribution proportional to the polarizability of each atom in the pair. In contrast, the TS-SCS method sometimes assigns negative polarizabilities to atoms-in-materials, and this makes subsequent calculations (e.g., MBD or van der Waals radius) crash.
- 3) M scaling provides a unified scaling law describing the different behaviors of isolated atoms and buried atoms. This allows MCLF to accurately describe both surface and buried atoms. The TS-SCS method (which does not use m scaling) could not accurately describe polar diatomic molecules irrespective of the partitioning method.
- 4) MCLF separates non-directional from directional screening of the dipole interaction tensor. The non-directionally screened polarizability is constrained to be less than or equal to the conduction limit upper bound, and this provides improved accuracy for buried atoms. In contrast, the TS-SCS method often produces atomic polarizabilities that are unphysically higher than the conduction limit.
- 5) MCLF uses a multibody screening function to capture the fluctuating dipole alignment at short distances and disorder at long distances. This computes more accurate C_6 coefficients.
- 6) MCLF computes three different types of screened dipole polarizabilities: (a) the non-directionally screened polarizabilities that are used as force-field and QDO input parameters, (b) the imfreq fluctuating polarizabilities that describe the local directional alignment contributing to C_6 coefficients, and (c) the static polarizability containing long-range directional alignment of dipoles due to a constant externally applied electric field. MCLF computes the full polarizability tensors including diagonal and off-diagonal components.

- 7) MCLF uses an iterative polarizability screening that improves accuracy and allows the computational cost to scale linearly with increasing number of atoms in the unit cell by avoiding matrix inversions. In contrast, the TS-SCS method uses matrix inversions that cause approximately cubic scaling of computational cost with increasing number of atoms in the unit cell. For this reason, MCLF is much more computationally efficient than TS-SCS.
- 8) MCLF parameterizes a Quantum Drude Oscillator (QDO) model to yield higher-order (e.g., quadrupolar and octupolar) atom-in-material polarizabilities and higher-order atom-in-material dispersion coefficients (e.g., C_8 , C_9 , C_{10}) and associated mixing rules.
- 9) The MCLF atom-in-material polarizability tensors are always symmetric, while the TS-SCS atom-in-material polarizability tensors are sometimes asymmetric. Symmetric polarizability tensors are more convenient, because they can be displayed as ellipsoids.

Tests were performed on diverse material types: isolated atoms, diatomic molecules, periodic solids, small organic and inorganic molecules, fullerenes, polyacenes, ice supercells containing up to 263424 atoms per unit cell, and an HIV reverse transcriptase biomolecule. For each test set in this study, MCLF gave $\leq 12\%$ MARE on the static polarizabilities and C_6 coefficients, and $\leq 14\%$ MARE for the static polarizability eigenvalues. This substantially improves over the TS-SCS method. For the static polarizabilities of solids: (a) TS-SCS with H, IH, and DDEC6 partitioning gave MARE of 47%, 30%, and 24%, respectively, (b) MCLF(DDEC6) gave MARE of 12%, and (c) all of the unscreened methods gave much larger errors than the screened methods. For the static polarizabilities of diatomic molecules: (a) TS-SCS(H) gave 41% MARE with a largest error of 437%, (b) TS-SCS(DDEC6) gave 37% MARE with a largest error of 440%, and (c) MCLF(DDEC6) gave 10% MARE with a largest error of 34%.

Conflicts of interest: There are no conflicts of interest to declare.

Acknowledgements:

This project was funded by National Science Foundation (NSF) CAREER Award DMR-1555376. Supercomputing resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE).¹⁵⁶ XSEDE is funded by NSF grant ACI-1053575. XSEDE project grant TG-CTS100027 provided allocations on the Comet clusters at the SDSC and the Stampede cluster at the TACC. The authors sincerely thank the technical support staff of XSEDE, TACC, and SDSC. Calculations on the HIV-RT complex were performed on the Rocket High Performance Computing Service at Newcastle University.

Electronic Supporting Information (ESI) available: A pdf file containing: damping radii definition (S1); def2QZVPPDD basis set definition (S2); m scaling formula for wp (S3); and plot of the smooth cutoff function (S4). A zip format archive containing: isolated atoms reference data and regression results; Rhombert integration and Richardson extrapolation coefficients; def2QZVPPDD.gbs and def2QZVPPDD.pseudo files containing the full basis set and RECP parameter tabulations, respectively; HIV reverse transcriptase complex geometry, ONETEP input files, and tabulated MCLF and TS-SCS results for every atom in the material; MCLF and TS-SCS input and output files for the CH_2Br_2 example showing symmetric (MCLF) and asymmetric (TS-SCS) atom-in-material polarizability tensors; $C_{8,A}$ and $C_{10,A}$ models for isolated atoms; and detailed MCLF and TS-SCS results for each material in the following test sets: diatomic molecules, small molecules and molecule pairs, solids, polyacenes and fullerenes.

8. References

1. L. P. Lee, D. J. Cole, C.-K. Skylaris, W. L. Jorgensen and M. C. Payne, *J. Chem. Theory Comput.*, 2013, **9**, 2981-2991.
2. L. P. Lee, N. G. Limas, D. J. Cole, M. C. Payne, C. K. Skylaris and T. A. Manz, *J. Chem. Theory Comput.*, 2014, **10**, 5377-5390.
3. D. J. Cole, J. Z. Vilseck, J. Tirado-Rives, M. C. Payne and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2016, **12**, 2312-2323.
4. P. Bleiziffer, K. Schaller and S. Riniker, *J. Chem. Inf. Model.*, 2018, **58**, 579-590.
5. A. Mondal and S. Balasubramanian, *J. Phys. Chem. B*, 2014, **118**, 3409-3422.
6. A. Mondal and S. Balasubramanian, *J. Phys. Chem. B*, 2015, **119**, 11041-11051.
7. A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield and G. J. Martyna, *Phys. Rev. B*, 2013, **87**, 144103.
8. A. J. Misquitta and A. J. Stone, *Arxiv preprint*, 2018, 1806.06737.
9. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
10. A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
11. K. Bica, M. Deetlefs, C. Schroder and K. R. Seddon, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2703-2711.
12. P. T. Kiss and A. Baranyai, *J. Chem. Phys.*, 2013, **138**, 204507.
13. P. Kiss and A. Baranyai, *J. Chem. Phys.*, 2012, **137**, 194103.
14. J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549-2564.
15. W. L. Jorgensen, K. P. Jensen and A. N. Alexandrova, *J. Chem. Theory Comput.*, 2007, **3**, 1987-1992.
16. W. L. Jorgensen, *J. Chem. Theory Comput.*, 2007, **3**, 1877-1877.
17. T. A. Halgren and W. Damm, *Curr. Opin. Struct. Biol.*, 2001, **11**, 236-242.
18. A. Warshel, M. Kato and A. V. Pisliakov, *J. Chem. Theory Comput.*, 2007, **3**, 2034-2045.
19. A. Ambrosetti, N. Ferri, R. A. DiStasio and A. Tkatchenko, *Science*, 2016, **351**, 1171-1176.
20. J. Hermann, R. DiStasio and A. Tkatchenko, *Chem. Rev.*, 2017, **117**, 4714-4758.
21. S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105-5154.
22. E. R. Johnson, I. D. Mackie and G. A. DiLabio, *J. Phys. Org. Chem.*, 2009, **22**, 1127-1135.
23. J. Klimes and A. Michaelides, *J. Chem. Phys.*, 2012, **137**, 120901.
24. R. DiStasio, V. Gobre and A. Tkatchenko, *J. Phys. Condens. Matter*, 2014, **26**, 213202.
25. G. Starkschall and R. G. Gordon, *J. Chem. Phys.*, 1972, **56**, 2801-2806.
26. J. Applequist, J. R. Carl and K. K. Fung, *J. Am. Chem. Soc.*, 1972, **94**, 2952-2960.
27. B. T. Thole, *Chem. Phys.*, 1981, **59**, 341-350.
28. A. Mayer and P. O. Astrand, *J. Phys. Chem. A*, 2008, **112**, 1277-1285.
29. A. Mayer, *Phys. Rev. B*, 2007, **75**, 045407.
30. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **122**, 154104.
31. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2005, **123**, 024101.
32. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2006, **124**, 014104.
33. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 154108.
34. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2006, **124**, 174104.
35. S. N. Steinmann and C. Corminboeuf, *J. Chem. Phys.*, 2011, **134**, 1-5.
36. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **123**, 154101.
37. F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129-138.
38. A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.

39. A. G. Donchev, *J. Chem. Phys.*, 2006, **125**, 074713.
40. T. Gould, *J. Chem. Phys.*, 2016, **145**, 084308.
41. E. R. Davidson and S. Chakravorty, *Theor. Chim. Acta*, 1992, **83**, 319-330.
42. P. Bultinck, C. Van Alsenoy, P. W. Ayers and R. Carbo-Dorca, *J. Chem. Phys.*, 2007, **126**, 144111.
43. A. V. Marenich, S. V. Jerome, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 527-541.
44. T. A. Manz and D. S. Sholl, *J. Chem. Theory Comput.*, 2012, **8**, 2844-2867.
45. T. Gould, S. Lebegue, J. G. Angyan and T. Bucko, *J. Chem. Theory Comput.*, 2016, **12**, 5920-5930.
46. T. Bucko, S. Lebegue, J. Hafner and J. G. Angyan, *J. Chem. Theory Comput.*, 2013, **9**, 4293-4299.
47. T. Bucko, S. Lebegue, J. Hafner and J. G. Angyan, *Phys. Rev. B*, 2013, **87**, 064110.
48. T. Bucko, S. Lebegue, J. G. Angyan and J. Hafner, *J. Chem. Phys.*, 2014, **141**, 034114.
49. A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 18A508.
50. R. E. Watson, *Phys. Rev.*, 1958, **111**, 1108-1110.
51. T. A. Manz and D. S. Sholl, *J. Chem. Theory Comput.*, 2010, **6**, 2455-2468.
52. T. A. Manz and N. Gabaldon Limas, *RSC Adv.*, 2016, **6**, 47771-47801.
53. T. A. Manz, *J. Comput. Chem.*, 2013, **34**, 418-421.
54. D. E. P. Vanpoucke, P. Bultinck and I. Van Driessche, *J. Comput. Chem.*, 2013, **34**, 405-417.
55. D. E. P. Vanpoucke, I. Van Driessche and P. Bultinck, *J. Comput. Chem.*, 2013, **34**, 422-427.
56. T. Gould and T. Bucko, *J. Chem. Theory Comput.*, 2016, **12**, 3603-3613.
57. N. Gabaldon Limas and T. A. Manz, *RSC Adv.*, 2016, **6**, 45727-45747.
58. T. A. Manz, *RSC Adv.*, 2017, **7**, 45552-45581.
59. N. Gabaldon Limas and T. A. Manz, *RSC Adv.*, 2018, **8**, 2678-2707.
60. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. J. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09, 2010, **Revision E.01**.
61. W. M. Haynes, ed., *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 2016-2017, 10.188-10.203.
62. Y. P. Kathuria, *Journal of Optics-Nouvelle Revue D Optique*, 1991, **22**, 149-151.
63. J. N. Wilson, *Chem. Rev.*, 1939, **25**, 377-406.
64. E. Talebian and M. Talebian, *Optik*, 2013, **124**, 2324-2326.
65. D. J. Margoliash and W. J. Meath, *J. Chem. Phys.*, 1978, **68**, 1426-1431.
66. G. D. Zeiss and W. J. Meath, *Mol. Phys.*, 1977, **33**, 1155-1176.
67. H. B. G. Casimir and D. Polder, *Phys. Rev.*, 1948, **73**, 360-372.
68. M. A. L. Marques, A. Castro, G. Mallocci, G. Mulas and S. Botti, *J. Chem. Phys.*, 2007, **127**, 014107.
69. A. Jiemchooraj, P. Norman and B. E. Sernelius, *J. Chem. Phys.*, 2005, **123**, 124312.

70. T. Gould and T. Bucko, *J. Chem. Theory Comput.*, 2016, **12**, 3603-3613.
71. K. T. Tang, *Phys. Rev.*, 1969, **177**, 108-114.
72. S. G. Porsev and A. Derevianko, *J. Chem. Phys.*, 2003, **119**, 844-850.
73. J. M. Tao, J. P. Perdew and A. Ruzsinszky, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 18-21.
74. J. G. McDaniel and J. R. Schmidt, *J. Phys. Chem. B*, 2014, **118**, 8042-8053.
75. K. T. Tang and J. P. Toennies, *J. Chem. Phys.*, 1984, **80**, 3726-3741.
76. M. Sadhukhan and F. R. Manby, *Phys. Rev. B*, 2016, **94**, 115106.
77. T. W. Whitfield and G. J. Martyna, *J. Chem. Phys.*, 2007, **126**, 074104.
78. A. Krawczuk-Pantula, D. Pérez, K. Stadnicka and P. Macchi, *Trans. Amer. Cryst. Ass.*, 2012, **42**, 1-25.
79. J. Applequist, *Acc. Chem. Res.*, 1977, **10**, 79-85.
80. P. Rostron, S. Gaber and D. Gaber, *Int. J. Eng. Tech. Res.*, 2016, **6**, 50-64.
81. R. S. Tobias, *J. Chem. Educ.*, 1967, **44**, 2-8.
82. W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, Addison-Wesley Publishing Company, Reading, MA, Second edn., 1962, 84-86.
83. T. Brinck, J. S. Murray and P. Politzer, *J. Chem. Phys.*, 1993, **98**, 4305-4306.
84. P. Y. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497-1506.
85. J. M. Wang, P. Cieplak, J. Li, T. J. Hou, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3091-3099.
86. J. M. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. J. Hsieh, H. X. Lei, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3100-3111.
87. D. Elking, T. Darden and R. J. Woods, *J. Comput. Chem.*, 2007, **28**, 1261-1274.
88. P. T. van Duijnen and M. Swart, *J. Phys. Chem. A*, 1998, **102**, 2399-2407.
89. J. G. Angyan, G. Jansen, M. Loos, C. Hattig and B. A. Hess, *Chem. Phys. Lett.*, 1994, **219**, 267-273.
90. A. Krishtal, P. Senet, M. Yang and C. Van Alsenoy, *J. Chem. Phys.*, 2006, **125**, 034312.
91. E. Heid, A. Szabadi and C. Schroder, *Phys. Chem. Chem. Phys.*, 2018, **20**, 10992-10996.
92. R. F. W. Bader, T. A. Keith, K. M. Gough and K. E. Laidig, *Mol. Phys.*, 1992, **75**, 1167-1189.
93. A. J. Stone, *Mol. Phys.*, 1985, **56**, 1065-1082.
94. B. A. Bauer, T. R. Lucas, A. Krishtal, C. Van Alsenoy and S. Patel, *J. Phys. Chem. A*, 2010, **114**, 8984-8992.
95. N. A. de Lima, *J. Chem. Phys.*, 2010, **132**, 014110.
96. E. B. Anders, *J. ACM*, 1966, **13**, 505-510.
97. J. Dutka, *Hist. Math.*, 1984, **11**, 3-21.
98. D. C. Joyce, *SIAM Rev.*, 1971, **13**, 435-488.
99. L. N. Trefethen and R. S. Schreiber, *Siam J. Matrix Anal. Appl.*, 1990, **11**, 335-360.
100. L. N. Trefethen, *ACM SIGNUM Newsletter*, 1985, **20**, 2-5.
101. V. Strassen, *Numerische Mathematik*, 1969, **13**, 354-356.
102. G. Schulz, *Zeitschrift für angewandte Mathematik und Mechanik*, 1933, **13**, 57-59.
103. E. W. Weisstein, Matrix Norm, <http://mathworld.wolfram.com/MatrixNorm.html>, (accessed August 2018, 2018).
104. A. Sivri, *Practical Extrapolation Methods: Theory and Applications*, Cambridge University Press, Cambridge, UK, 1 edn., 2003.
105. M. Hellenbrandt, *Crystallogr. Rev.*, 2004, **10**, 17-22.
106. G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169-11186.
107. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865-3868.
108. A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, *Acta Cryst. B*, 2002, **58**, 364-369.
109. W. A. Saidi and P. Norman, *J. Chem. Phys.*, 2016, **145**, 024311.

110. J. M. Tao, Y. X. Mo, G. C. Tian and A. Ruzsinszky, *Phys. Rev. B*, 2016, **94**, 085126.
111. J. Kauczor, P. Norman and W. A. Saidi, *J. Chem. Phys.*, 2013, **138**, 114107.
112. I. G. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*, John Wiley & Sons, West Sussex, England, 2006, 183-254.
113. Q. Yang, D. Liu, C. Zhong and J.-R. Li, *Chem. Rev.*, 2013, **113**, 8261-8323.
114. R. O. Dror, R. M. Dirks, J. P. Grossman, H. F. Xu, D. E. Shaw and D. C. Rees, *Annu. Rev. Biophys.*, 2012, **41**, 429-452.
115. D. Dubbeldam, A. Torres-Knoop and K. S. Walton, *Mol. Simul.*, 2013, **39**, 1253-1292.
116. D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81-101.
117. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236.
118. B. Chen and J. I. Siepmann, *J. Phys. Chem. B*, 1999, **103**, 5370-5379.
119. Y. Zhang, A. Otani and E. J. Maginn, *J. Chem. Theory Comput.*, 2015, **11**, 3537-3546.
120. K. Vanommeslaeghe and A. D. MacKerell, *J. Chem. Inf. Model.*, 2012, **52**, 3144-3154.
121. B. L. Bush and R. P. Sheridan, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 756-762.
122. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474-6487.
123. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor, *J. Phys. Chem. B*, 2010, **114**, 7830-7843.
124. C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.*, 2014, **10**, 865-879.
125. P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik and J. Vondrasek, *J. Comput. Chem.*, 1997, **18**, 1136-1150.
126. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kucsera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1614.
127. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Comput. Mol. Sci.*, 2013, **3**, 198-210.
128. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
129. S. L. James, *Chem. Soc. Rev.*, 2003, **32**, 276-288.
130. H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
131. Y. J. Cui, Y. F. Yue, G. D. Qian and B. L. Chen, *Chem. Rev.*, 2012, **112**, 1126-1162.
132. Q. Xu and C. L. Zhong, *J. Phys. Chem. C*, 2010, **114**, 5035-5042.
133. S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy and R. Schmid, *Phys. Status Solidi B*, 2013, **250**, 1128-1141.
134. L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier and V. Van Speybroeck, *J. Comput. Chem.*, 2015, **36**, 1015-1027.
135. S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schutt and K. R. Muller, *Sci. Adv.*, 2017, **3**, e1603015.
136. V. Botu, R. Batra, J. Chapman and R. Ramprasad, *J. Phys. Chem. C*, 2017, **121**, 511-522.
137. I. Leontyev and A. Stuchebrukhov, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2613-2626.
138. J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, *Chem. Rev.*, 2016, **116**, 4983-5013.
139. Y. Shi, Z. Xia, J. J. Zhang, R. Best, C. J. Wu, J. W. Ponder and P. Y. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046-4063.
140. O. Borodin, *J. Phys. Chem. B*, 2009, **113**, 11463-11478.

141. R. Kurth and N. Bannert, eds., *Retroviruses: Molecular Biology, Genomics and Pathogenesis* Caister Academic Press, Norfolk, UK, 2010.
142. M. Bollini, R. Gallardo-Macias, K. A. Spasov, J. Tirado-Rives, K. S. Anderson and W. L. Jorgensen, *Bioorganic Med. Chem. Lett.*, 2013, **23**, 1110-1113.
143. D. J. Cole, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2014, **10**, 565-571.
144. D. J. Cole, J. Tirado-Rives and W. L. Jorgensen, *Biochim. Biophys. Acta-General Subjects*, 2015, **1850**, 966-971.
145. C. K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Chem. Phys.*, 2005, **122**, 084119.
146. N. D. M. Hine, P. D. Haynes, A. A. Mostofi, C. K. Skylaris and M. C. Payne, *Comput. Phys. Commun.*, 2009, **180**, 1041-1053.
147. K. Das, A. D. Clark, P. J. Lewi, J. Heeres, M. R. de Jonge, L. M. H. Koymans, H. M. Vinkers, F. Daeyaert, D. W. Ludovici, M. J. Kukla, B. De Corte, R. W. Kavash, C. Y. Ho, H. Ye, M. A. Lichtenstein, K. Andries, R. Pauwels, M. P. de Bethune, P. L. Boyer, P. Clark, S. H. Hughes, P. A. J. Janssen and E. Arnold, *J. Med. Chem.*, 2004, **47**, 2550-2560.
148. W. L. Jorgensen and J. Tirado-Rives, *J. Comput. Chem.*, 2005, **26**, 1689-1700.
149. W. L. Jorgensen, *Acc. Chem. Res.*, 2009, **42**, 724-733.
150. L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.*, 1982, **48**, 1425-1428.
151. A. Ruiz-Serrano, N. D. M. Hine and C. K. Skylaris, *J. Chem. Phys.*, 2012, **136**, 234101.
152. A. A. Mostofi, P. D. Haynes, C. K. Skylaris and M. C. Payne, *J. Chem. Phys.*, 2003, **119**, 8842-8848.
153. J. C. Womack, L. Anton, J. Dziedzic, P. J. Hasnip, M. I. J. Probert and C. K. Skylaris, *J. Chem. Theory Comput.*, 2018, **14**, 1412-1432.
154. M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499-3509.
155. E. T. Walters, M. Mohebifar, E. R. Johnson and C. N. Rowley, *J. Phys. Chem. B*, 2018, **122**, 6690-6701.
156. J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkins-Diehr, *Comput. Sci. Eng.*, 2014, **16**, 62-74.