JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

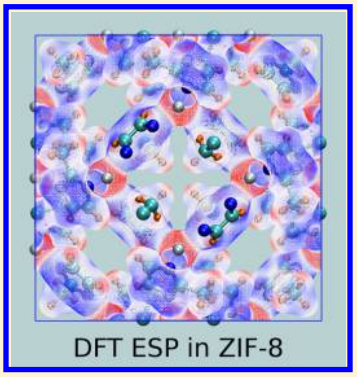
Partial Charges in Periodic Systems: Improving Electrostatic Potential (ESP) Fitting via Total Dipole Fluctuations and Multiframe Approaches

Andrea Gabrieli, Marco Sant,* Pierfranco Demontis, and Giuseppe B. Suffritti

Dipartimento di Chimica e Farmacia, Università degli Studi di Sassari, Via Vienna 2, 07100 Sassari, Italy

Supporting Information

ABSTRACT: Two major improvements to the state-of-the-art Repeating Electrostatic Potential Extracted Atomic (REPEAT) method, for generating accurate partial charges for molecular simulations of periodic structures, are here developed. The first, D-REPEAT, consists in the simultaneous fit of the electrostatic potential (ESP), together with the total dipole fluctuations (TDF) of the framework. The second, M-REPEAT, allows the fit of multiple ESP configurations at once. When both techniques are fused into one, DM-REPEAT method, the resulting charges become remarkably stable over a large set of fitting regions, giving a robust and physically sound solution to the buried atoms problem. The method capabilities are extensively studied in ZIF-8 framework, and subsequently applied to IRMOF-1 and ITQ-29 crystal structures. To our knowledge, this is the first time that this approach is proposed in the context of periodic systems.



1. INTRODUCTION

Microporous materials, because of their peculiar topology and favorable surface-to-volume ratio, have found broad application in the chemical industry, primarily as catalysts and molecular sieves.[1−3] Among these, metal−organic frameworks (MOFs),[4,5] have recently attracted the attention of the scientific community. This is due to their remarkable properties, such as chemical and thermal stability, tunable pore size and topology, and large surface area, which exceeds that of other sorbents.[6] For these reasons, besides being employed in traditional applications, MOFs are currently the subject of extensive research as promising materials for environmental applications (e.g., hydrogen storage[7−12] and carbon capture[13−16]) and biomedical applications (e.g., drug delivery[17−23]).

Adsorption and diffusion properties of guest molecules within these materials are routinely studied by classical molecular simulations such as molecular dynamics (MD) and Grand Canonical Monte Carlo (GCMC).[24−27] In fact, these techniques are able to follow the evolution of the system over large time and space scales, not accessible to *ab initio* methods. However, these computations require a force field in order to be performed (with the overall accuracy depending on its quality). Many types of force fields exist;[28−34] the large majority of them rely on a fixed point charge description to model the electrostatic interactions, according to the Coulomb law,

$$E = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

where q are the charges of atoms i and j, which are separated by a distance r, and ϵ_0 is the vacuum permittivity. This choice is particularly appealing, because of its low computational burden and its ability to give direct insight on the chemical properties of the system. Still, there is a problem: partial charges are not observable quantities and cannot be directly measured. Therefore, in the past 30 years, many routes have been proposed to generate reliable partial charges, spanning from empirical to *ab initio* approaches.[35−63]

Quantum mechanical (QM) methods are nowadays widely adopted, because of their accuracy in describing the system charge distribution and the increase in computational power available. In particular, one can fit the charges to reproduce the electrostatic potential (ESP) over a large number of grid points around a molecule, lying outside its van der Waals (vdW) surface; this approach has proved to give good results for a large number of systems.[61,64]

In the context of microporous crystals further care should be taken: in fact, in periodic systems, the ESP surface is only defined to within an arbitrary additive constant. The Repeating Electrostatic Potential Extracted Atomic (REPEAT) charges method, developed by Campañá and co-workers,[65] has successfully tackled the problem and has been widely adopted.[61]

The flaw of ESP methods is that statistical underdetermination of some charges may arise in specific systems, leading to unphysical values. This occurs when some atoms are buried within a shell of other atoms (e.g., tetrahedral carbons): in this case, the ESP points used in the fit are all farther from the buried atom than from the surrounding atoms. This results in a

Received: March 2, 2015
Published: July 10, 2015

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

small influence of the buried atom over the fit. Exploiting this fact, the most common way to overcome this problem consists of restraining the buried atom charges to a reasonable value,[44,65,66] with a negligible deterioration of the fit quality. Nonetheless, this approach of biasing the fit toward a target value introduces some degree of arbitrariness in the choice of the target itself.

Some alternative methods to the use of restrained ESP fitting, which are not affected by the presence of buried atoms, have been proposed. The dipole preserving and polarization consistent (DPPC) charges method,[54] based on a Lagrangian optimization to reproduce atomic dipoles and multipoles,[67] has been developed with the introduction of peculiar features, including charge redistribution based on relative atomic electronegativity and the treatment of molecular dipole polarization. Resulting charges are able to accurately reproduce the molecular dipole moments, thus ensuring a good description of the system local properties (which may be spoiled by basic ESP fitting).

Another approach, based on the atoms in molecules[68] (AIM) theory, is the density-derived electrostatic and chemical (DDEC)[50−52] method. Here, the electron density is partitioned and assigned to each atom employing spherically symmetrical weighting functions,[69] with the partial charges optimized to simultaneously reproduce realistic chemical states, together with the QM electrostatic potential. This method can be applied to a large variety of systems and has been proven to give good results in MOFs, which are also consistent across an ensemble of system configurations generated using *ab initio* molecular dynamics[70] (AIMD).

A different approach, which is not based on QM computations, is the charge equalization (Qeq) method, originally developed for molecules by Rappé and Goddard,[43] and subsequently extended to MOFs.[71−74] Recently, two further improvements have been presented,[62] namely, FC-Qeq and I-Qeq, which perform better than previous rapid charge assignment methods and are suitable for fast screening of microporous materials.

In this contribution, we will discuss the case of Zeolitic Imidazolate Framework 8 (ZIF-8), where the problem of buried atoms is severe.[51] A physically sound method to unambiguously determine the charge values is proposed. It is based on two independent key features. The first is the simultaneous fit of both the ESP and the fluctuations of the total dipole moment (TDF) about its mean value. The second is the simultaneous fit of multiple ESP configurations at once. This approach, being an extension of the REPEAT method, will be named DM-REPEAT, from the fusion of the two methods: the D-REPEAT (total Dipole fluctuations REPEAT) and the M-REPEAT (Multiframe REPEAT). To our knowledge, this is the first time that this approach has been proposed for microporous crystals.

The paper is structured as follows: in Section 2, we present the theoretical background of the method and its implementation, while in Section 3, the method capabilities are explored by studying three microporous materials of great practical interest (namely, ZIF-8, IRMOF-1, and ITQ-29). These frameworks are chosen, in particular, because of their high symmetry and the presence of buried atoms, which makes them a good test for the method proposed here. Finally, some conclusions are given.

2. COMPUTATIONAL METHODS

2.1. Theoretical Background. *ESP Fitting.* ESP-based methods consist of determining the set of partial charges that best reproduces a reference electrostatic potential. This reference is computed on a mesh via first-principles methods, and the fit is performed over a given ensemble of grid points,[44] discarding those lying within the vdW surface.[39] The problem can be formulated as the solution of an overdetermined system of linear equations.

In a periodic crystal of N atoms, the model ESP $\tilde{E}_i = \tilde{E}(\mathbf{r}_i)$ for the ith grid point located at $\mathbf{r}_i$ can be computed according to the Ewald formalism:[25]

$$\tilde{E}_i = \sum_{\mathbf{n}} \sum_j q_j \frac{\mathrm{erfc}(\sqrt{\alpha}|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} + \sum_{\mathbf{k}\neq 0} \sum_j q_j \left(\frac{4\pi}{Vk^2}\right) e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} e^{[-k^2/(4\alpha)]} \quad (1)$$

where q_j is the partial charge of the jth crystal atom located in $\mathbf{r}_j$, and V is the simulation box volume. This formula is extensively discussed elsewhere;[75−77] for the sake of convenience, we remember that $\sqrt{2/\alpha}$ is the width of the Gaussian compensating charge distribution, erfc(·) is the complementary error function, $\mathbf{n}$ represents the vectors associated with the periodic box replica, and $\mathbf{k}$ represents the reciprocal space vectors.

Focusing on a single grid point i, the problem of determining the parameters that minimize the error e_i in

$$E_i = \tilde{E}_i + e_i \quad (2)$$

where E_i is the reference ESP value.

As can be noted in eq 1, the model ESP is linearly dependent on the fitting parameters q_j. Exploiting this property, eq 2 can be rewritten as

$$E_i = \sum_j \frac{\partial \tilde{E}_i}{\partial q_j} q_j + e_i \quad (3)$$

Now considering the entire ensemble of G grid points, we end up with G equations of the form of eq 2. The resulting system can be written in matrix form as follows:

$$\mathbf{B} = \mathbf{Aq} + \mathbf{e} \quad (4)$$

where $\mathbf{A}$ is a (G,N) matrix of elements $A_{ij} = \partial \tilde{E}_i/\partial q_j$; $\mathbf{B}$ and $\mathbf{e}$ are vectors of elements E_i and e_i respectively; and $\mathbf{q}$ is the vector of the N model parameters q_j.

The fit consists of solving a least-squares problem, minimizing the following figure of merit:

$$\chi = \sum_{i=1}^{G} e_i^2 = \sum_{i=1}^{G} (E_i - \tilde{E}_i)^2 = \|\mathbf{B} - \mathbf{Aq}\|^2 \quad (5)$$

The minimization is achieved when

$$\frac{\partial \chi}{\partial q_j} = 0 \qquad \forall\, j \quad (6)$$

It is known from linear algebra that eq 6 leads to the following system:[78]

$$(\mathbf{A}^T\mathbf{A})\mathbf{q} = \mathbf{A}^T\mathbf{B} \quad (7)$$

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

where $\mathbf{A}^T\mathbf{A}$ is a symmetric $N \times N$ matrix. The number of equations, then, is greatly reduced, moving from the number of grid points G (of the order of 10^6) to the number of unknowns N (of the order of 10^2).

In the case of periodic crystals, in accordance with the REPEAT approach, the fit should be performed over the relative differences of the potential for both reference and model ESP. The figure of merit here becomes

$$\chi = \sum_{i=1}^{G} [(E_i - \langle E \rangle) - (\tilde{E}_i - \langle \tilde{E} \rangle)]^2 \tag{8}$$

where the quantities within angle brackets are averaged over the full set of grid points. Writing this in matrix form, in analogy with eq 5, the elements of $\mathbf{B}$ and $\mathbf{A}$ in $\|\mathbf{B} - \mathbf{Aq}\|^2$ are now

$$B_i = E_i - \langle E \rangle$$
$$A_{ij} = \frac{\partial \tilde{E}_i}{\partial q_j} - \langle A_j \rangle \tag{9}$$

with $\langle A_j \rangle = (1/G)\sum_i \partial \tilde{E}_i / \partial q_j$.

It is also necessary to ensure that the total model charge equals the total charge of the reference system, q_{tot}, which should be zero for periodic structures. This is achieved by introducing a constraint via the Lagrange multiplier λ.[42] The merit function now reads

$$\chi = \|\mathbf{B} - \mathbf{Aq}\|^2 + \lambda\left(\sum_{j=1}^{N} q_j - q_{tot}\right) \tag{10}$$

The system of equations to be solved is once again obtained by zeroing the derivatives of the merit function, in this case: $\partial\chi/\partial q_j = 0$ for all q_j and $\partial\chi/\partial\lambda = 0$. With respect to eq 7, the system is now given as

$$\begin{cases} \mathbf{A}^T\mathbf{Aq} + \mathbf{\Lambda q} = \mathbf{A}^T\mathbf{B} \\ \sum_{j=1}^{N} q_j = q_{tot} \end{cases} \tag{11}$$

where vector $\mathbf{\Lambda}$ is composed of N elements λ (where a 0.5 factor in front of λ has been dropped, given that the latter is an arbitrary constant).[42] In matrix notation, eq 11 becomes

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} & 1 \\ a_{21} & a_{22} & \cdots & a_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \lambda \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \\ q_{tot} \end{bmatrix} \tag{12}$$

where a_{ij} are simply the elements of the matrix $\mathbf{A}^T\mathbf{A}$, and b_i are the elements of the vector $\mathbf{A}^T\mathbf{B}$, with $\mathbf{A}$ and $\mathbf{B}$ being defined in eqs 9.

If symmetrically unique atoms are present in the crystal, they should have the same charge. As noted elsewhere,[44] and as will be extensively discussed later, the approach of averaging together the charge of similar atoms after the fit leads to unpredictable effects on its quality. For this reason, the symmetry should be enforced during the fit. This can be achieved by summing together the columns of matrix $\mathbf{A}$ corresponding to symmetric atoms. This way, T groups made of symmetric unique atoms are formed, each bearing a different charge, thus reducing the total number of parameters from N to T. The second equation in eq 11 then becomes

$$\sum_{j=1}^{T} t_j q_j = q_{tot} \tag{13}$$

where t_j is the number of symmetrically equivalent atoms associated with the jth fit parameter and the following holds: $\sum_j^T t_j = N$. Finally, the "1" values appearing in the last row and last column of the left-hand side matrix of eq 12 are now replaced by the corresponding t_j values. For the reasons extensively discussed in Section 3.1.3, unless stated otherwise, throughout this work, the atom-type symmetry will be enforced.

TDF Fit. The electric dipole moment is an essential quantity in chemistry and physics, giving a measure of the separation of charges in the system and describing the response of the system to an external electric field.[79,80]

Being interested in periodic structures, the attention is here focused on the total dipole fluctuations (TDF) of the entire unit cell. The problem in this case consists of finding the set of partial charges giving the classical model TDF that best reproduces the reference TDF. The latter is computed via the Berry phase approach,[81,82] by performing an *ab initio* molecular dynamics simulation. At the same time, the model total dipole moment $\tilde{\mathbf{p}}$ for a given configuration of the N crystal atoms is

$$\tilde{\mathbf{p}} = \sum_{j}^{N} q_j \mathbf{r}_j \tag{14}$$

where q_j is the charge associated with the jth atom located at $\mathbf{r}_j$. Here, the reference point is chosen to be the origin of coordinates. Note that, in a neutral system, $\tilde{\mathbf{p}}$ is independent of the choice of the reference point.

We consider a trajectory of the system, made by S frames. In a periodic crystal, the total dipole moment is defined up to a constant term. It is then useful to study the fluctuations of this quantity about its mean value (TDF). Given the close analogy to the ESP fitting problem, see eq 8, we can write the merit function as

$$\chi = \sum_{i=1}^{S} [(p_i - \langle p \rangle) - (\tilde{p}_i - \langle \tilde{p} \rangle)]^2 \tag{15}$$

where p_i is the ith reference total dipole moment for each Cartesian component, and the angle brackets represent the averages over the full trajectory. Writing this in matrix form, $\|\mathbf{B} - \mathbf{Aq}\|^2$, the elements of $\mathbf{B}$ and $\mathbf{A}$ are

$$B_i = p_i - \langle p \rangle$$
$$A_{ij} = \frac{\partial \tilde{p}_i}{\partial q_j} - \langle A_j \rangle \tag{16}$$

with $\langle A_j \rangle = (1/S)\sum_i \partial \tilde{p}_i / \partial q_j$. Here, $\partial \tilde{p}_i / \partial q_j$ is equal to r_{ij}, which is the coordinate (for each component direction) of the jth atom in the ith frame.

Finally, considering the three Cartesian components together, we can write

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}_x^{\mathrm{T}}\mathbf{A}_x + \mathbf{A}_y^{\mathrm{T}}\mathbf{A}_y + \mathbf{A}_z^{\mathrm{T}}\mathbf{A}_z$$
$$\mathbf{A}^{\mathrm{T}}\mathbf{B} = \mathbf{A}_x^{\mathrm{T}}\mathbf{B}_x + \mathbf{A}_y^{\mathrm{T}}\mathbf{B}_y + \mathbf{A}_z^{\mathrm{T}}\mathbf{B}_z \quad (17)$$

From here on, the fit proceeds exactly as in the ESP case described by eqs 10−13. If needed, the constraint of the total charge of the system and the treatment of the symmetrically equivalent atoms can be applied.

Simultaneous ESP and TDF Fit. It is now possible to fit simultaneously the ESP of C system configurations together with the TDF along the three coordinates of a system trajectory (not necessarily made by the C configurations). As in eqs 17, the solution comes from adding the $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ matrices of ESP and TDF together:

$$\begin{aligned}\mathbf{A}^{\mathrm{T}}\mathbf{A} &= \frac{(1-w)^2}{M_{\mathrm{ESP}}}(\mathbf{A}^{\mathrm{T}}\mathbf{A})_{\mathrm{ESP}} + \frac{w^2}{M_{\mathrm{TDF}}}(\mathbf{A}^{\mathrm{T}}\mathbf{A})_{\mathrm{TDF}} \\ &= \frac{(1-w)^2}{M_{\mathrm{ESP}}}\sum_{c=1}^{C}\mathbf{A}_c^{\mathrm{T}}\mathbf{A}_c + \frac{w^2}{M_{\mathrm{TDF}}}\sum_{d=1}^{3}\mathbf{A}_d^{\mathrm{T}}\mathbf{A}_d \\ \mathbf{A}^{\mathrm{T}}\mathbf{B} &= \frac{(1-w)^2}{M_{\mathrm{ESP}}}(\mathbf{A}^{\mathrm{T}}\mathbf{B})_{\mathrm{ESP}} + \frac{w^2}{M_{\mathrm{TDF}}}(\mathbf{A}^{\mathrm{T}}\mathbf{B})_{\mathrm{TDF}} \\ &= \frac{(1-w)^2}{M_{\mathrm{ESP}}}\sum_{c=1}^{C}\mathbf{A}_c^{\mathrm{T}}\mathbf{B}_c + \frac{w^2}{M_{\mathrm{TDF}}}\sum_{d=1}^{3}\mathbf{A}_d^{\mathrm{T}}\mathbf{B}_d\end{aligned} \quad (18)$$

Here, special care has been taken in balancing the influence of each contribution according to its total magnitude: $M_{\mathrm{ESP}} = \sum_c M_c$, where the magnitude of the cth configuration is $M_c = \mathbf{B}_c^{\mathrm{T}}\mathbf{B}_c = \sum_i (B_i)^2$ with B_i as defined in eqs 9, and $M_{\mathrm{TDF}} = \sum_d M_d$, where $M_d = \mathbf{B}_d^{\mathrm{T}}\mathbf{B}_d = \sum_i (B_i)^2$, with B_i as defined in eqs 16. Moreover, a weighting factor w has been introduced to control the balance between ESP and TDF: this, for example, allows one to switch between a pure ESP ($w = 0$) and pure TDF ($w = 1$) fit. Again, the system total charge constraint and the symmetric equivalent atoms are treated as previously explained.

2.2. Implementation. *Reference Data Computations.* The reference data for the fit are obtained by performing Born−Oppenheimer Molecular Dynamics (BOMD) simulations and single-point calculations relying on the CP2K open source code.[83−87] Density functional theory (DFT)[88,89] computations are performed to evaluate the system energy according to the Gaussian and Plane Waves (GPW)[90] method.

The system structure is generated[91] starting from experimental data. The simulation box coincides with one unit cell of the crystal. After thermalizing at 300 K, the system trajectory is computed in the canonical NVT ensemble (weak coupling with canonical sampling through velocity-rescaling CSVR thermostat[92]) with a time step of 0.5 fs. To achieve the best compromise between accuracy and computational cost for the investigated systems, Perdew−Burke−Ernzerhof (PBE)[93] functional, along with Goedecker−Teter−Hutter (GTH) pseudopotential,[94−96] and GTH basis sets[84] have been used. The basis sets for each type of atom are triple-ζ with two sets of polarization functions (TZV2P), except for Zn (TZVP), in accordance with previous works.[97−99] The DFT-D3 method[100] is used to take into account dispersion interactions. The charge density cutoff is 700 Ry (for ITQ-29 only, this value is increased to 1000 Ry). The system is fully periodic. The convergence criterion for the SCF (largest element of the wave function gradient) is 10^{-7}.

The total dipole moments $\mathbf{p}_i$ are computed at each BOMD step via the Berry phase approach;[82] these are used to generate the B_i elements of eq 16 for the TDF part of the fit. The thus-computed total dipole moment is defined[80] modulo the lattice periodicity $\mathbf{a}$ (i.e., branching). This may lead to discontinuities in the TDF signal, which can be removed by refolding all total dipole moments into the same branch (e.g., the branch of the first one). This is achieved by using a procedure analogous to the minimum image approach,[24] with the $\mathbf{p}_{i+1}$ total dipole value obtained as follows:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \Delta\mathbf{p} - \mathbf{a}\left[\mathrm{nint}\left(\frac{\Delta\mathbf{p}}{\mathbf{a}}\right)\right] \quad (19)$$

where $\Delta\mathbf{p} = \mathbf{p}_{i+1} - \mathbf{p}_i$ and nint$(\cdot)$ is the function returning the nearest integer. Clearly, care should be taken in using consecutive total dipole values that are close enough to ensure that all jumps larger than one branch are artifacts to be removed.

The ESP values (E_i) are computed over a mesh with a spacing of ∼0.2 Å, and written in a Gaussian cube file for selected system snapshots. This spacing guarantees that the number of grid points is sufficient to ensure a well-converged fit[65,66] (i.e., doubling the number of points does not alter the fit results). The ESP values are then used to obtain the B_i elements of eqs 9.

Model Data Computations. The procedure to compute the A_{ij} elements of eqs 16 for the TDF part of the fit is straightforward: these are simply obtained by taking the $\partial\tilde{p}_i/\partial q_j$ values from the BOMD trajectory ($\partial\tilde{p}_i/\partial q_j = r_{ij}$). Regarding the model ESP, instead, the A_{ij} elements of eqs 9 are computed with a modified version of the MUSIC code.[101] This gives the electrostatic potential felt at grid points having the same position as in the reference cube file due to the interaction with each system atom. Throughout this work, the parameter α of the Ewald summation is set to 0.156 Å^{-2}. This choice ensures the convergence of the real space part of eq 1 within one unit cell; the sum is then truncated at $\mathbf{n} = 0$ and the minimum image convention is applied.[24] The number of reciprocal space vectors spans from −15 to 15 (neglecting very small terms by setting the MUSIC locut parameter to 10^{-10}). In particular, the $\partial\tilde{E}_i/\partial q_j$ values (see eqs 9) are obtained by setting the charge of the jth atom (or of the symmetrically equivalent atoms belonging to the jth fit parameter) to unity and the charges of the remaining atoms to zero. This procedure is closely related to the one followed in our previous work[99] to generate model data for the optimization of bonded interactions via the *force matching* method.

Fitting Code. In order to solve the linear system of equations, and obtain the optimal parameters, we developed a Python program,[102,103] which will be freely available from the group institutional web page.

The quality of the fitted parameters is assessed by computing, independently for ESP and TDF, the root of the relative mean square error ($\mathrm{RRMS_E}$ and $\mathrm{RRMS_T}$, respectively). From algebraical manipulation of $\|\mathbf{B} - \mathbf{Aq}\|^2 = (\mathbf{B} - \mathbf{Aq})^{\mathrm{T}}(\mathbf{B} - \mathbf{Aq})$ (see eq 5), this reads

$$\mathrm{RRMS_S} = \sqrt{\frac{\mathbf{B}^{\mathrm{T}}\mathbf{B} - 2\,\mathbf{q}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{B} + \mathbf{q}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Aq}}{\mathbf{B}^{\mathrm{T}}\mathbf{B}}} \quad (20)$$

where $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \sum_s \mathbf{B}_s^{\mathrm{T}}\mathbf{B}_s$ with the s index running either on ESP ($\mathrm{RRMS_E}$) configurations or TDF ($\mathrm{RRMS_T}$) directions used in the fit (see eqs 18). The same holds for $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

A key point of this implementation is the automatic choice of the optimal weight w (see eqs 18), which controls the relative importance of the ESP and TDF parts in the minimized merit function. We recall that fitting algorithms yield the set of parameters that give the best numerical match between model and reference data, sometimes spoiling the underlying physics just to attain a negligible improvement in the final value of the merit function. Clearly, the addition of any extra contribution (e.g., TDF) to the merit function will deteriorate, to a certain extent, the quality ($RRMS_E$) of the ESP part, with respect to a pure ESP fit. The goal of a simultaneous match, then, is to keep the deterioration of the ESP quality to a minimum, while improving the TDF one. This is achieved through an algorithm that first computes and stores the $RRMS_P$ for both pure ESP and pure TDF. Then, a loop is performed by varying at each ith iteration the value of w, ranging from 0.0 to 1.0 in small steps. Starting from the second iteration, the relative change in RRMS, with respect to the pure value, is computed for both ESP and TDF, as follows:

$$\Delta R = \frac{\mathrm{RRMS}_i - \mathrm{RRMS}_P}{\mathrm{RRMS}_P}$$

The last weight for which $\Delta R_{ESP} < \Delta R_{TDF}$ holds is stored (w_A), just before the degradation in the ESP fit quality becomes greater than the potential improvement in the TDF one. At the same time, to ensure that $RRMS_E$ at the stopping weight is not larger than $RRMS_T$, a second weight (w_B) is stored at the minimum absolute difference between ESP and TDF RRMS:

$$\min\{\mathrm{abs}(\mathrm{RRMS}_{E,i} - \mathrm{RRMS}_{T,i})\}$$

The optimum result is the one obtained at whichever of the two conditions occurs earlier ($\min\{w_A,w_B\}$, i.e., smaller TDF weight). It is important to note that the computational burden of this automatic procedure is negligible because, after generating the initial arrays of data, only a matter of seconds is required to solve the linear system multiple times for the various weights.

Finally, we point out that the practice of accumulating all data in the $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{B}$ matrices here adopted (eqs 18) is particularly effective for the simultaneous fit of several ESP configurations. In fact, in this context, the system of equations to be solved has a fixed size N (see eq 7), independent of the number of snapshots treated. This is not the case if one uses the solver directly over eq 4, where the number of equations increases linearly with the number of configurations. Since the data of each snapshot must be processed explicitly, the problem becomes quickly unmanageable.

The code has been validated against a reference cube file generated from a classical system of known charges. The code has been proven to be capable to recover the charges within an accuracy of 10^{-4}.

3. RESULTS AND DISCUSSIONS

3.1. ZIF-8. Zeolitic imidazolate frameworks (ZIFs),[104−106] which are members of the MOF family, are composed of clusters of transition metal bridged by imidazole organic linkers, mimicking the tetrahedral Si−O arrangement found in zeolites.

In particular, ZIF-8 (see Figure 1) is currently the subject of intensive research,[23,107−117] primarily in the context of gas separation applications. This material has a cubic unit cell with lattice parameters of $a = b = c = 16.9856$ Å, and it is comprised of Zn metal ions with 2-methylimidazole (MeIM) as a ligand.

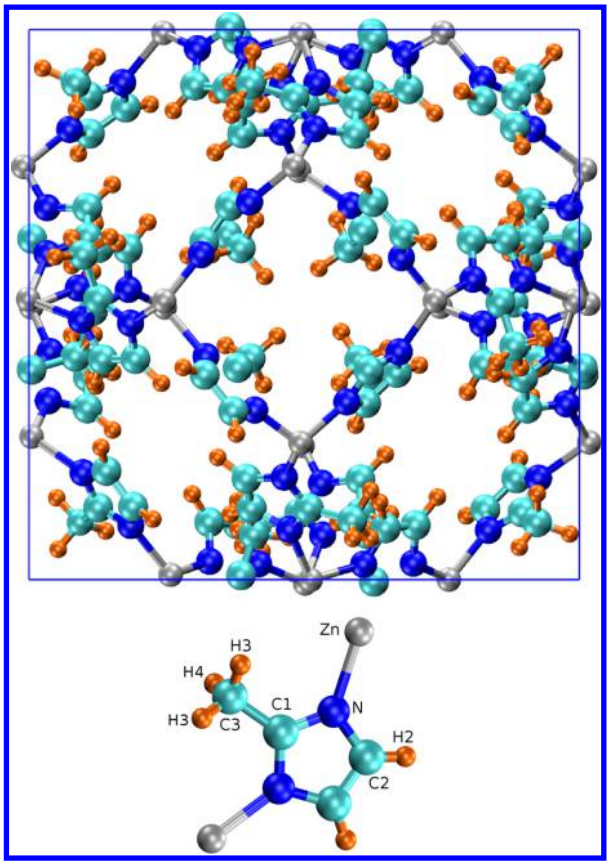


Figure 1. ZIF-8 framework. [Legend: cyan, C; orange, H; blue, N; and gray, Zn; the accompanying numbers indicate the various symmetric unique types.]

One unit cell contains 276 atoms, and these can be grouped into seven different types, according to their symmetry: 24 C1, 48 C2, 24 C3, 48 H2, 72 H3, 48 N, and 12 Zn. Generally, the methyl hydrogens (H3) should all bear the same charge. However, in the crystallographic configuration, the H3 are split in two subgroups (i.e., 48 H3 and 24 H4), because of their peculiar space arrangement. This distinction disappears spontaneously in any other frame taken from a BOMD trajectory.

3.1.1. Crystallographic REPEAT. Standard REPEAT fitting is here applied to the crystallographic configuration of ZIF-8[107] (giving $C = 1$ in eqs 18) and a thorough analysis of the results and related problems is carried out. Note that, throughout this work, the REPEAT fits are performed without restraints; at the same time, all reported charge values are in units of the elementary charge (e). The strong dependence of the fitted partial charges on the size of the exclusion region, made of vdW spheres, is discussed. In particular, this size is varied by multiplying the vdW radius of each crystal atom (values taken from the UFF database[30]) by a scaling factor γ, ranging from 0.6 to 2.5.

In a previous work of our group, the partial charges for ZIF-8 have been derived with the REPEAT method.[98] The value of γ was set to 1.0, giving charges in agreement with fragment computations.

As first noted by Watanabe and co-workers,[51] in ZIF-8, there is a strong dependence of the charges on γ, with the Zn charge spanning from positive values (0.69 for $\gamma = 1$) to negative values (−2.58 for $\gamma = 2$).

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

To further investigate this behavior, a deeper (e.g., more γ values) analysis is performed here with our new code implementation. The dramatic instability of the fitted partial charges, as a function of γ, is shown in Figure 2a. All atoms, except the hydrogens, widely fluctuate, which makes the extrapolation of a reliable value impracticable. The latter fact can be appreciated in Figure 3. Here, the QM reference ESP is compared with those generated by using partial charges fitted at $\gamma = 1.0$ and $\gamma = 1.4$. Although the $\gamma = 1.4$ charges perform well in the region used for their fit, they fail outside, giving unphysical values, see Table 1. On the other hand, the $\gamma = 1.0$ charges have chemically intuitive values, and the resulting ESP is in fair agreement with the QM reference also in the region excluded from the fit.

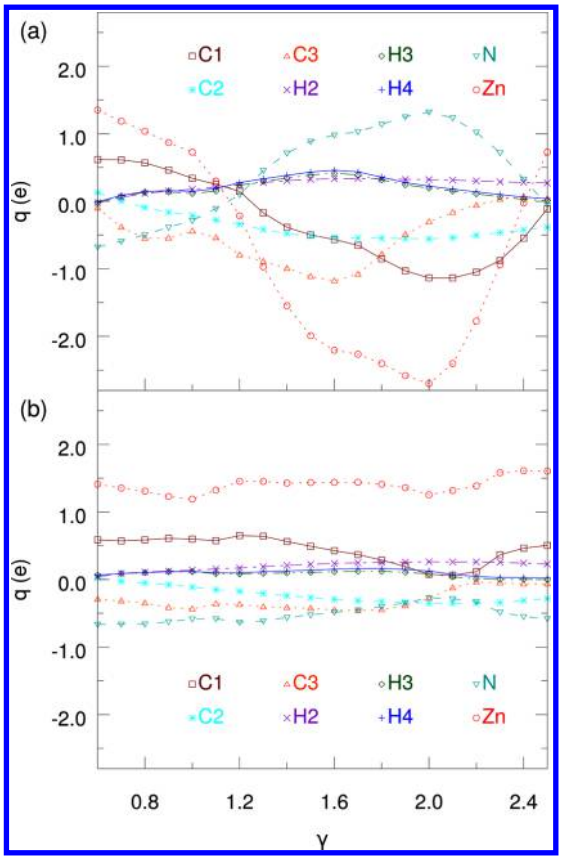


Figure 2. (a) REPEAT vs (b) D-REPEAT partial charges for crystallographic ZIF-8, as a function of the increasing scaling factor γ for the vdW radius.

Optimum γ. At this point, it is worthwhile to briefly discuss the choice of γ. Close to the nuclei, the point charge model fails (i.e., bad $RRMS_E$) while moving farther away, the representation improves. This behavior can be appreciated considering two simple molecules, namely, CO_2 and CH_4. In Figure 4 the partial charges for these molecules are reported as a function of γ (computations are made with the same setup used for the crystals, placing a single molecule in the center of a periodic box having same lattice parameters as ZIF-8). It is clear how the results attain a stable value only at $\gamma > 1.2$ for CO_2 and $\gamma > 1.5$ for CH_4.

In dealing with microporous materials, one is interested in properly reproducing the interactions of the framework with some given sorbate. Then, the focus is on the ESP region corresponding to the sorbate accessible volume. The analysis of pure CO_2 and CH_4 BOMD trajectories in ZIF-8 (unit cell loaded with 12 and 20 molecules, respectively) shows that this region is never closer than $\gamma = 1.2$. This is in agreement with the value of $\gamma = 1.3$ used by Chen and co-workers.[66] Clearly, results from a fit over this region can be trusted only when they are stable, as in Figure 4 and not as in Figure 2a.

On the other hand, an ESP fit performed only over grid points at large γ may blur the detail of atoms that are close together (e.g., random charge of buried atoms). In MD simulations of sorbate molecules having size comparable to that of the host pores, reliable dynamical properties can be obtained only by using a flexible framework.[118] In this case, the crystal intramolecular interactions should be properly reproduced as well, and the partial charge detail lost at large γ is no more negligible. As an example, the charges resulting from the fits at $\gamma = 1.0$ and $\gamma = 1.4$ are reported in Table 1. Here, note that the $RRMS_E$ value for $\gamma = 1.0$ is twice that at $\gamma = 1.4$, giving the wrong impression of a greater quality of the latter. This $RRMS_E$ improvement, as already mentioned, comes from the fact that the point charge model is more effective at large γ. Actually, the numerical values for $\gamma = 1.4$ lack the detail of the buried atoms, as previously shown in Figure 3. This would ultimately lead to a poor reproduction of the intramolecular interactions.[44]

3.1.2. Crystallographic D-REPEAT. The total dipole fluctuations (TDF) are now taken into account. In Figure 5, a representative extract of the TDF signal coming from a trajectory consisting of 10^4 BOMD steps is reported, together with the classical TDF obtained over the same trajectory by

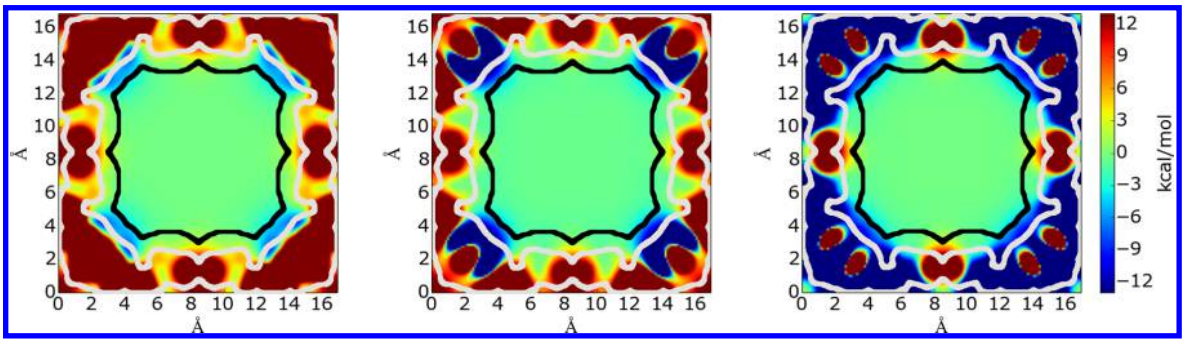


Figure 3. ESP contour plots for crystallographic ZIF-8: *yz*-plane located at *x* equal to half box size. QM reference data (left), data derived from charges fitted via standard REPEAT at $\gamma = 1.0$ (center) and $\gamma = 1.4$ (right). The borders of the fitting regions are depicted by contour lines; focusing on the central cage, the first line (black) delimits the $\gamma = 1.4$ region, while the second line (gray) delimits the $\gamma = 1.0$ region. The third outermost line (gray) indicates the beginning of the $\gamma = 1.0$ region for the adjacent cages.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

Table 1. Partial Charges for ZIF-8, Shown Via a Comparison of Different Methods: REPEAT (R, C = 1 and S = 0 in eqs 18), Total Dipole Fluctuations (TDF, C = 0 and S = 10^4), D-REPEAT (D-R, C = 1 and S = 10^4, M-REPEAT (M-R, C = 100 and S = 0), and DM-REPEAT (DM-R, C = 100 and S = 10^4) for Various Configurations and γ Values[a]

type	R 1.0γ[b]	R 1.4γ[b]	TDF	D-R 1.4γ[b]	M-R 1.4γ	DM-R 1.4γ	DDEC(m)[c]	CBAC[d]
C1	0.3407	−0.3859	0.2996	0.5626	0.2827	0.3586	0.4291	0.6460
C2	−0.2109	−0.4749	−0.0656	−0.2400	−0.1950	−0.1641	−0.0839	−0.0620
C3	−0.4421	−0.9927	−0.0533	−0.4142	−0.1889	−0.2174	−0.4526	−0.0410
H2	0.1774	0.3046	0.1573	0.2095	0.1652	0.1617	0.1128	0.0100
H3	0.1143	0.3421	0.0710	0.0981	0.0615	0.0692	0.1325	0.0100
H4	0.1471	0.3829	0.0696	0.1277	0.0615	0.0694	0.1306	0.0100
N	−0.2857	0.7130	−0.7403	−0.5623	−0.2484	−0.4533	−0.3879	−0.4400
Zn	0.7282	−1.5478	1.6786	1.4266	0.5562	1.1248	0.6920	0.6970
$RRMS_E$	0.2965	0.1382		0.1935	0.3772	0.4909		
$RRMS_T$			0.3572	0.5057		0.4920		
$EVAL_E$	0.4931	1.5369	0.9930	0.6862	0.3772	0.4909	0.4784	1.8587
$EVAL_T$	0.6917	1.7539	0.3572	0.5057	0.7004	0.4920	0.6443	0.7519

[a]For comparison with the literature, the last two columns report values from the DDEC(m)[51] and CBAC[51,53] methods. EVAL is the RRMS evaluated imposing the specific set of charges over C = 100 or S = 10^4; see text. [b]ESP fit over crystallographic configuration. [c]Data taken from ref 51. [d]Values taken from ref 51, according to method described in ref 53.

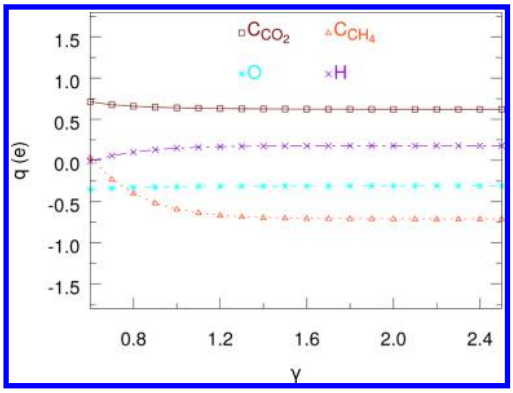


Figure 4. REPEAT partial charges for CO_2 and CH_4, as a function of γ.

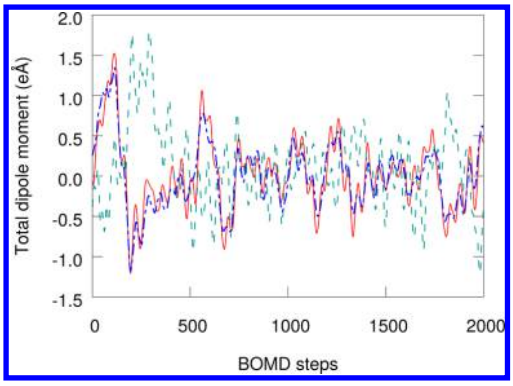


Figure 5. Total dipole signal from QM reference data (red continuous line), pure TDF fitted charges (blue dots and dashes), and REPEAT charges at γ = 1.4 (turquoise dashes).

using the ESP charges previously fitted at γ = 1.4, and the classical TDF obtained with charges generated through the direct fit of the reference TDF signal (corresponding values are reported in Table 1). The latter signal agrees well with the QM reference one, while the ESP derived signal performs much worse, missing even the qualitative features of the reference.

Concerning the trajectory length required for a reliable fit of the TDF, we have performed an analysis by computing the mean absolute percentage deviation,

$$\mathrm{MAPD} = \frac{100}{T}\sum_{i=1}^{T}\left|\frac{(q_i^{\mathrm{ref}} - q_i)}{q_i^{\mathrm{ref}}}\right| \qquad (21)$$

between a reference set of T partial charges, q_i^{ref}, obtained from a TDF fit over 30 000 consecutive BOMD steps (15 ps) and a set, q_i, obtained over a shorter trajectory. For a 5000 frames (2.5 ps) fit, the MAPD is close to 10%, while for 10^4 frames (5 ps), this value decreases to 5%, with the latter giving a good compromise between accuracy and computational cost.

For the above reasons, all TDF fits throughout this paper, unless otherwise stated, are performed over 10^4 frames (giving S = 10^4 in eqs 18). Here, the charges are generated by matching the crystallographic ESP and the trajectory TDF (i.e., D-REPEAT) simultaneously, over 20 different γ values, ranging from 0.6 to 2.5. For each γ, 101 fits are performed by changing the weight of the dipole part over the merit function, ranging from 0.0 (pure ESP) to 1.0 (pure TDF). Note that the entire procedure, thanks to the current implementation, requires just minutes to accomplish. To elucidate how the optimal weight is automatically chosen, in Figure 6, the ESP and TDF RRMS are plotted as a function of the TDF percentile weight at γ = 1.4.

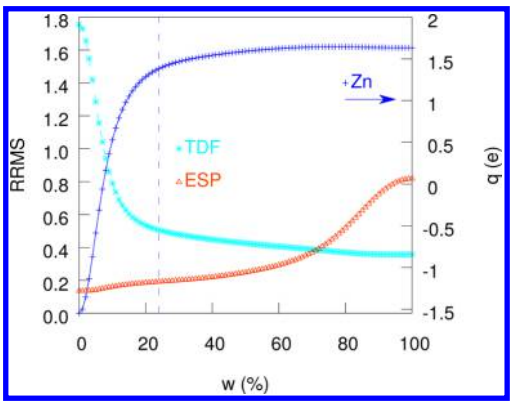


Figure 6. RRMS evolution for ESP and TDF, D-REPEAT in ZIF-8, as a function of the TDF percentile weight over the fit (values shown on the left y-axis). The vertical line is located at the stopping value (24%). The trend in the Zn partial charge is also reported (values shown on the right y-axis).

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

The vertical line is located at the stopping weight value (w = 0.24): here, the relative degradation of the ESP, with respect to its best, becomes larger than the TDF counterpart (see Section 2.2). The trend of the Zn partial charge is also reported, with corresponding values on the right y-axis. It is easy to see that a negative charge value of the Zn atom (being the most unstable among others, see Figure 2a) leads to a completely incorrect TDF. In any case, poorly determined charges can be recovered with a small weight of the TDF part: this is sufficient to guide the fit toward more-reasonable values with limited degradation of the ESP $RRMS_E$.

In Figure 2b, the charges obtained at the optimal weight for all γ are depicted. It is easy to see the neat stability improvement, with respect to Figure 2a, where no TDF is used. More importantly, the obtained results attain chemically intuitive values over a wide range of γ.

3.1.3. DM-REPEAT. As seen in the previous sections, it is common practice to generate the partial charges by taking into account only the crystallographic structure. A fit performed over this perfectly symmetric configuration may spoil the resulting partial charges, leading to unphysical results. This is due to the fact that the useful information is actually limited to the information carried by the asymmetric unit. Moreover, a good classical MD representation of the system cannot be achieved without relaxing the experimental structure (i.e., without taking into account the framework flexibility, see Section 3.1.1). In this case, a full force field parametrization is required, implying that a large amount of QM data has been previously collected.[99] It is then advantageous to improve the statistics by performing the simultaneous fit of a set of system configurations.

In Figure 7a, charges coming from M-REPEAT are shown, where the ESP reference data are made by 100 frames obtained by striding the BOMD trajectory every 100 steps (i.e., sampling the trajectory every 50 fs ensures the statistical independence of the various configurations). The good convergence of the resulting charges has been assessed computing the MAPD (see eq 21), with respect to a reference set obtained from a M-REPEAT fit over 100 configurations generated by striding a trajectory of 30000 BOMD steps every 300 frames (i.e., every 150 fs). The charges are converged within 3%.

In comparison with Figure 2a, the multiframe approach clearly stabilizes the results at large γ. Note that, in a multiframe fit, as expected, the methyl H atoms (here, H3 and H4) attain the same charge value spontaneously (see Table 1).

From previous considerations, we know that the results of a fit at large γ lack some details regarding the buried atoms. This can also be seen, for example, considering the trend of the results at γ < 1.2 in Figure 7a. In order to overcome this problem, the DM-REPEAT method should be employed. In Figure 7b, the corresponding partial charges are shown. The stability over the full γ range is remarkable, with values that are also physically sound. In view of the work done, numerical results for γ = 1.4 (see Table 1) should be taken as the optimum charges: the chosen value of γ ensures that ESP data are fitted over a region corresponding to the accessible volume for a large number of sorbates, the multiframe approach guarantees that this is true for all system configurations, while the TDF part addresses the lost details related to the ESP statistically underdetermined atoms (i.e., buried atoms).

As opposed to crystallographic ZIF-8, the DM-REPEAT fit reaches the stopping criterion at the crossing point between ESP and TDF RRMSs, to prevent the $RRMS_E$ from becoming

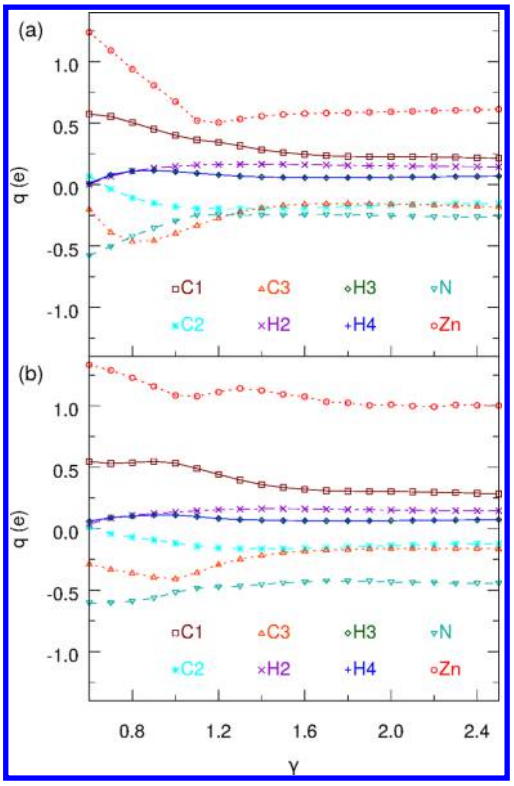


Figure 7. (a) M-REPEAT vs (b) DM-REPEAT partial charges for ZIF-8, as a function of increasing γ. Simultaneous ESP fit performed over 100 statistically independent frames.

larger than the TDF one. This different behavior is due to the fact that, in the multiframe approach, the TDF starts from a better RRMS than in the crystallographic case, 0.70 vs 1.75, respectively, while the starting ESP value is worse, 0.38 vs 0.14. This fact postpones the point where the relative degradation of ESP with respect to its best becomes larger than the TDF one, which, in this case, occurs after the crossing of the two curves. A representative example at γ = 1.4 is given in Figure 8.

Given that the RRMSs are related only to the set of data used for the specific fit, these values can be misleading, when comparing charges obtained from different sets. For this purpose, then, an evaluation index (EVAL) has been introduced. This is simply the RRMS computed from eq 20 by imposing a given set of charges q over the larger set of data

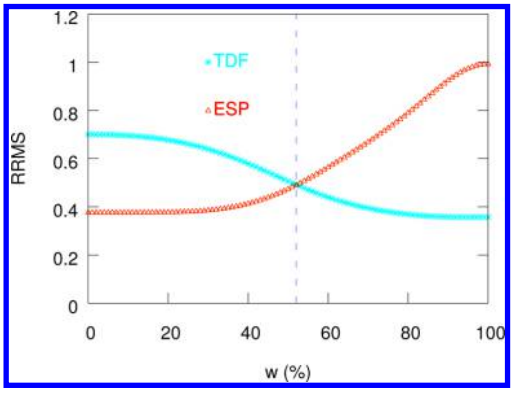


Figure 8. RRMS evolution for ESP and TDF, DM-REPEAT in ZIF-8, as a function of the TDF percentile weight over the fit. The vertical line is located at the stopping value (52%).

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

available (i.e., the DM-REPEAT one): C = 100 configurations for ESP ($EVAL_E$), and $S = 10^4$ BOMD trajectory frames for TDF ($EVAL_T$). For example, in Table 1, the $RRMS_E$ for R 1.4γ, performed over the crystallographic configuration, is far smaller (0.1382) than the corresponding M-R value (0.3772), performed over 100 independent configurations, giving the wrong impression that the first is better than the latter. Actually, looking at the corresponding $EVAL_E$ values (1.5369 vs 0.3772), the higher accuracy of the latter set of charges becomes clear. From an analysis of the EVALs parameters in Table1, it can be clearly appreciated that the DM-REPEAT set of charges gives the best compromise in accurately reproducing both ESP and TDF.

For comparison purposes, Table 1 also reports two sets of partial charges taken from the literature (see the last two columns). The first set, obtained through the DDEC(m)[50,51] method, reproduces the ESP very well, with an $EVAL_E$ (0.4784) even better than DM-REPEAT (0.4909). The TDF, instead, is less accurate, with a worse $EVAL_T$ (0.6443 vs 0.4920), most probably due to fact that the TDF is not explicitly taken into account by DDEC (in analogy to the M-REPEAT, which has a comparable $EVAL_T$ of 0.7004). The second set of charges has been obtained with the CBAC method, which is based on the assumption that the charges are dependent on the atoms bonding connectivity and that, in different MOFs, atoms with same connectivity should have identical charges.[53] The reported values are taken from ref 51, where the cell neutralization has been performed. In this case, the ESP reproduction is poor ($EVAL_E$ = 1.8587), most probably because these charges have been derived from numerous fragments, aiming at transferability, while the peculiar structure of ZIF-8 requires a tailored approach. The TDF, instead, is consistent with the other methods that do not explicitly include the fit of this property.

3.1.3.1. Importance of Atom-Type Symmetry and Multiframe in ESP Fitting. For clarity, this discussion is focused on the partial charges of the 12 Zn atoms in ZIF-8; still, its considerations are completely general. From the asymmetric unit, we know that these 12 atoms are symmetrically unique. In the crystallographic configuration, this results in all of the Zn atoms actually being replicas of only one atom and, thus, carrying less information than what may appear. On the other hand, in a given configuration (e.g., of the BOMD trajectory), these Zn atoms experience a heterogeneous environment; thus, the fit, in principle, can assign a different charge to each one of them. Such an environment becomes statistically equivalent when many frames are considered; for this reason, they should have the same partial charge. As a consequence, in a simultaneous fit of many frames, this property emerges spontaneously (see the last column of Table 2), where the partial charges of the 12 Zn atoms have been individually fitted over 100 statistically independent frames (i.e., one every 50 fs) for γ = 1.4. These results must be compared with the other columns showing the same fit over 1 and 10 frames, which are the first and the first 10 of the set of 100 frames, respectively. In particular, the standard deviation of the charges decreases with the number of frames used, while the mean converges toward a well-defined value. Note how the charge value of every single Zn atom in the last column is already converged, because of the simultaneous multiframe fit. The same charge smoothing effect can be achieved with less frames by imposing the atom-type symmetry during the fit (see the "sym" row in Table 2), where the unique value for all Zn atoms approaches that obtained over 100 frames with as few as 10 frames (provided that these also are statistically independent). Looking at the "1 fr" column in Table 2, it is evident that simply averaging the charges of the 12 Zn atoms together, without the use of atom-type symmetry, leads to unacceptable results (−0.0982). Note that the spread of charges is related to the buried nature of the Zn atoms and arises at γ > 1.0; this is due to the loss of detail in the description of the framework buried atoms occurring at large γ. The same spread is also responsible for the degradation in $RRMS_E$ when the atom-type symmetry is imposed (see Table 2). Over a single given configuration, in fact, better $RRMS_E$ values can be achieved if one leaves all the Zn free to attain a different charge, rather than imposing the atom-type symmetry. The better $RRMS_E$ is misleading, since it does not corresponds to a real improvement in the quality of the partial charges. These charges, in fact, perform well only on one frame but poorly on all the others (which conflicts with the basic requirements of a fixed point charge model). This can be quantitatively appreciated by looking at the $EVAL_E$ values (measure of charges quality over the largest set of data) along the columns in Table 2. Note that, in the crystallographic configuration, the fictitious $RRMS_E$ improvement occurs even when the atom-type symmetry is imposed since all atoms of a given type are, by construction, actually the same.

The effectiveness of atom-type symmetry in stabilizing the results comes from the fact that, in a given configuration, some of the Zn atoms are more statistically determined than others (because of their more favorable position, i.e., nearer to the crystal surface), thus leading the fit toward better charges. This causes the degradation of the $RRMS_E$, when imposing the symmetry, since all Zn atoms are forced to have the same charge.

Table 2. Individual Zn Partial Charges, Along with Their Mean, Standard Deviation, and $RRMS_E$ for a Pure ESP Fit over 1, 10, and 100 Frames (γ = 1.4) Taken from the Same Set (See Text)[a]

Zn	1 fr	10 fr	100 fr
1	−0.9031	0.6300	0.6140
2	−0.0683	0.4243	0.5785
3	−0.3897	0.6487	0.5564
4	−0.6014	0.7843	0.5822
5	0.4456	0.7905	0.6036
6	0.6184	0.3775	0.5762
7	0.6015	0.4406	0.5992
8	1.1697	0.5743	0.5205
9	0.6222	0.6998	0.5163
10	−0.9006	0.6530	0.5536
11	−1.0414	0.8162	0.5714
12	−0.7309	0.5594	0.5814
mean	−0.0982	0.6166	0.5711
std	0.7596	0.1470	0.0303
$RRMS_E$	0.1171	0.2910	0.3613
$EVAL_E$	1.6808	0.5640	0.3613
sym	0.3736	0.5460	0.5562
$RRMS_E$	0.4110	0.3924	0.3772
$EVAL_E$	0.4261	0.3802	0.3772

[a]In the last three rows, values attained by imposing atom-type symmetry are also reported. The $RRMS_E$ is computed over each individual subset used for the fit. $EVAL_E$ is the RRMS evaluated by imposing the specific set of charges over C = 100.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

The same considerations given above, showing the advantages of imposing the atom-type symmetry, rather than averaging the symmetric unique charges after the fit, can be applied also to the multiframe approach. This, in fact, should be preferred with respect to the averaging of various single configuration fits. In this case, the $RRMS_E$ degradation, when moving from single to multiframe approach, appears only if the atom-type symmetry is not imposed, thus evidencing the spurious nature of this effect. When the atom-type symmetry is imposed, in fact, moving across the columns of Table 2, a decreasing $RRMS_E$ trend is found, in agreement with the $EVAL_E$ trend.

3.2. IRMOF-1. This material, also known as MOF-5,[119] is part of the MOF family and has received much attention from the scientific community[51,65,66] as a starting point toward the rational design of novel materials with tunable porosity for efficient hydrogen storage applications.[7,9,120]

IRMOF-1 crystals are made by ZnO_4 tetrahedra (see Figure 9), joined together by 1,4-benzenedicarboxylate (BDC) groups,

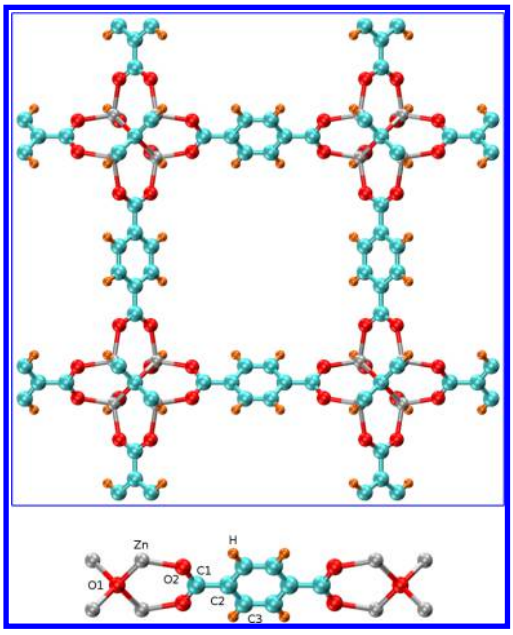


Figure 9. IRMOF-1 framework. [Legend: cyan, C; orange, H; red, O; gray, Zn; the accompanying numbers indicate the various symmetric unique types.]

arranged in a cubic structure (space group $Fm\bar{3}m$) having lattice parameters[119] of $a = b = c = 25.6690$ Å. The framework unit cell is composed of 424 atoms, which can be grouped into 7 types, according to their symmetry: 48 C1, 48 C2, 96 C3, 96 H, 8 O1, 96 O2, and 32 Zn.

The dependence of partial charges from γ are depicted in Figure 10. In Figure 10a, the REPEAT method is applied to the crystallographic structure, showing a certain degree of instability (even if less severe than in ZIF-8), mainly for the O1 atom type. The numerical values for $\gamma = 1.0$ are reported in Table 3, showing good agreement with the computations performed by Campañá and co-workers.[65] InFigure 10b, the M-REPEAT approach is applied over 100 frames, obtained by striding a 8000-step BOMD trajectory (only in this case, with DFT-relaxed cell parameters $a = b = c = 26.1371$ Å). The stability is good but the charges (see, as a representative example, the values at $\gamma = 1.4$ reported in Table 3) are quite

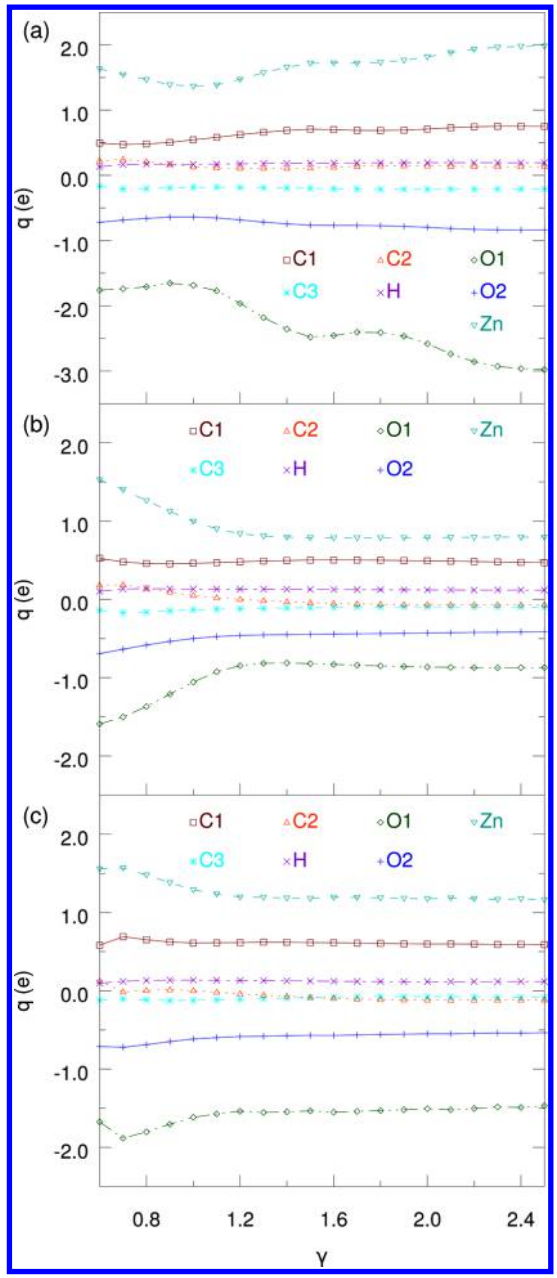


Figure 10. Partial charges for IRMOF-1 as a function of increasing γ: (a) REPEAT over crystallographic configuration, (b) M-REPEAT, and (c) DM-REPEAT.

small, giving a bad $EVAL_T$ for TDF. If the dipole fluctuations also are taken into account, for DM-REPEAT, shown in Figure 10c, the charges stability is preserved, while the reproduction of the TDF is improved, as can be seen in the corresponding column of Table 3. For the choice of γ, all considerations made for ZIF-8 hold, with $\gamma = 1.4$ giving optimum charges.

Also reported in Table 3 are two sets of charges taken from the literature, obtained with the DDEC/c3[52] and CBAC[51,53] methods. The first one is the evolution of the DDEC(m)[50,51] method. Here, it has been preferred, since it gives a better reproduction of *ab initio* ESP (see the entry for "IRMOF-1(PW91)" in Table 8 of ref 52). When compared with DM-REPEAT, the two methods have similar accuracy, with DM-REPEAT performing better, as can be seen from both the $EVAL_E$ and $EVAL_T$ values. Concerning CBAC charges, here

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

Table 3. Partial Charges for IRMOF-1: REPEAT (R, C = 1 and S = 0 in eq 18), M-REPEAT (M-R, C = 100 and S = 0), and DM-REPEAT (DM-R, C = 100 and S = 8000) Fits at Various γ[a]

type	[b]R 1.0γ	M-R 1.4γ	DM-R 1.4γ	[c]DDEC/c3	[d]CBAC
C1	0.5442	0.4971	0.6193	0.7046	0.8095
C3	−0.1831	−0.1099	−0.0939	−0.0755	−0.1265
C2	0.1303	−0.0297	−0.0692	−0.0920	0.0535
H	0.1705	0.1288	0.1269	0.1198	0.1125
O1	−1.6850	−0.8132	−1.5458	−1.2672	−1.9175
O2	−0.6389	−0.4498	−0.5743	−0.6182	−0.7895
Zn	1.3640	0.7949	1.1852	1.1196	1.5955
$RRMS_E$	0.1731	0.2880	0.3504		
$RRMS_T$			0.4960		
$EVAL_E$	0.4772	0.2880	0.3504	0.4220	0.9174
$EVAL_T$	0.4709	0.6252	0.4960	0.5144	0.4334

[a]For comparison, the last two columns report values from DDEC/c3[52] and CBAC[51,53] methods. EVAL is the RRMS-specific set of charges over C = 100 or S = 8000. [b]ESP fit over crystallographic configuration. [c]Data taken from ref 52. [d]Values from ref 51. according to the method described in ref 53.

they perform better than in ZIF-8, yet the ESP is poorly reproduced. Even though the quality of TDF is better than that of DM-REPEAT ($EVAL_T$ 0.4334 vs 0.4960), the latter clearly gives a better compromise in reproducing both ESP and TDF.

3.3. ITQ-29. The last material investigated here is ITQ-29,[121] the purely siliceous form of zeolite A;[122] it is extensively studied for its molecular sieving properties.[1,27,123,124] The cubic unit cell has lattice parameters[122] $a = b = c$ = 11.9190 Å. The framework is made by 72 atoms arranged as corner-sharing SiO_4 tetrahedra, see Figure 11. They are grouped in four types, according to symmetry, as follows: 12 O1, 24 O2, 12 O3, and 24 Si.

In Figure 12, the behavior of the partial charges is shown as a function of increasing γ. Looking at Figure 12a, the instability of the results, when REPEAT is applied to the crystallographic configuration, is evident. For comparison purposes, numerical

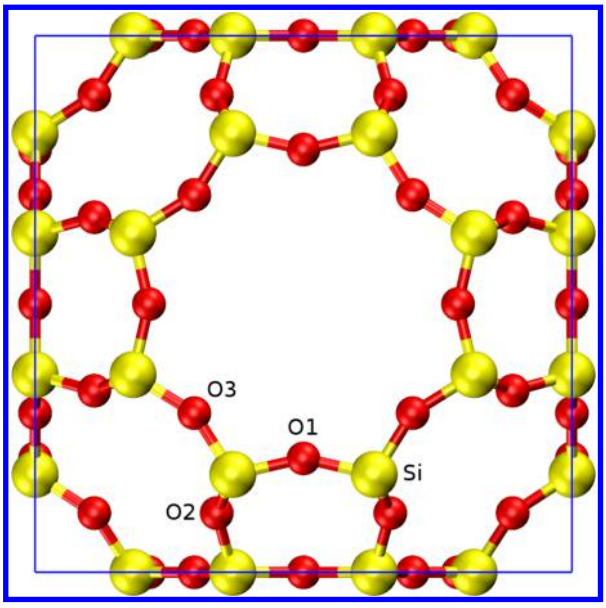


Figure 11. ITQ-29 framework. [Legend: yellow, Si; red, O; the accompanying numbers indicate the various symmetric unique types.]

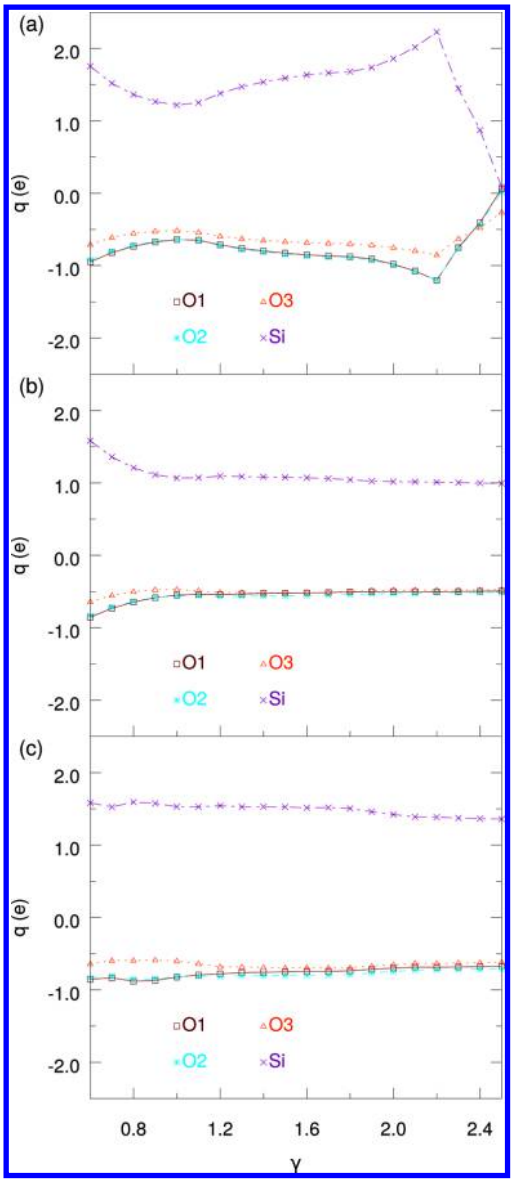


Figure 12. Partial charges for ITQ-29, as a function of increasing γ: (a) REPEAT over crystallographic configuration, (b) M-REPEAT, and (c) DM-REPEAT.

values at γ = 1.4 are reported in Table 4. In Figure 12b, the stability improves thanks to the M-REPEAT approach, applied over 100 frames obtained by striding a 10^4-step BOMD trajectory. Still, the magnitude of the charges (see Table 4) seems to be smaller than those expected from X-ray diffraction (XRD) experiments,[127] and from the rationale that the Si−O bond is half ionic and half covalent in character, leading to halved formal charges.[126] Chemically intuitive values are recovered with the DM-REPEAT approach (see Table 4), while the stability is preserved. This can be appreciated in Figure 12c. Once again, the optimum charges are associated with a value of γ = 1.4. Note that all the O atom types (O1, O2, O3) approach the same value when multiple frames are fitted simultaneously.

In Table 4, charge values from the literature are also reported. The first set is generated through the fitting of experimental isotherms.[125] The reproduction of ESP is good, very close to DM-REPEAT, but the one of TDF is poor. This comes from the fact that total dipole fluctuations were not

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

Table 4. Partial Charges for ITQ-29: REPEAT (R, C = 1 and S = 0 in eq 18), M-REPEAT (M-R, C = 100 and S = 0), and DM-REPEAT (DM-R, C = 100 and S = 10^4) Fits at γ = 1.4[a]

type	[b]R 1.4γ	M-R 1.4γ	DM-R 1.4γ	from ref 125	from ref 126
O1	−0.7968	−0.5265	−0.7570	−0.3930	−1.0000
O2	−0.8118	−0.5607	−0.8044	−0.3930	−1.0000
O3	−0.6508	−0.5169	−0.6910	−0.3930	−1.0000
Si	1.5356	1.0824	1.5284	0.7860	2.0000
$RRMS_E$	0.0721	0.2501	0.3925		
$RRMS_T$			0.4232		
$EVAL_E$	0.4272	0.2501	0.3925	0.4057	1.7050
$EVAL_T$	0.4210	0.5803	0.4232	0.6928	0.3001

[a]For comparison, the last two columns report values obtained from the fitting of experimental isotherms[125] and from heuristic reasoning[126] (see text). EVAL is the RRMS evaluated by imposing the specific set of charges over C = 100 or S = 10^4. [b]ESP fit over crystallographic configuration.

taken into account in the method used for the generation of charges. The second set is obtained by halving the atoms formal charges,[126] based on the previously mentioned chemical reasoning that Si−O bonds are half ionic and half covalent in character, in agreement with XRD data.[127] Even though the $EVAL_T$ is lower than all the others reported in the table, the $EVAL_E$ is far too large, showing a reproduction of TDF that is too unbalanced, at the expense of ESP.

3.4. Further Considerations. The method here presented is considered to be part of the generation of a full force field (bonded plus nonbonded interactions), tailored to a specific structure to ensure accurate computations. In this context, the collection of a large amount of QM data becomes cost-effective, since the data are used multiple times to optimize the parameters for the various types of interactions.[99]

In cases where only partial charges are sought (e.g., when the bonded interactions are taken from a database or a fixed framework model is employed), as suggested by one of the referees, it is possible to reduce the time required to obtain the QM reference data by avoiding the generation of a full BOMD trajectory. The DFT computations, in fact, can be performed over a set of uncorrelated configurations extracted from a classical trajectory coming from an MD run using a crude force field.

This option has been tested for ZIF-8 by employing a force field made only by bonded interactions (i.e., having no long-range dispersion and electrostatic terms) with the relevant parameters taken from ref 111. A 1 ns trajectory is simulated with NAMD[128] in the microcanonical (NVE) ensemble at T = 300 K with a time step of 1 fs. Subsequently, 100 snapshots (one every 10 ps) are extracted from this trajectory and used to compute the *ab initio* ESP and total dipole with CP2K. Fitting these data, the resulting charges fall within an accuracy of 7% (MAPD, see eq 21), with respect to the DM-REPEAT ones obtained over the BOMD trajectory. At the same time, the $EVAL_E$ (0.5039) and $EVAL_T$ (0.4905) are in very good agreement with the more computationally expensive ones. All these observations support the feasibility of the fast approach presented here.

In what follows, we briefly report some options that have been tested and discarded during this work. A large number of possible routes, to reach a stable solution in the generation of partial charges, may be explored. However, many of these are not useful.

The first one consists of fitting the electric field instead of the ESP. This quantity, being the spatial derivative of the ESP along the three Cartesian components, suffers the same problems of ESP fitting,[129] with the drawback of a computational burden that is three times larger.

The second option is to fit only ESP grid points that are not included within a region containing a given percentage of QM electron density (e.g., an increasing value starting from 99.5%). For a comparable number of grid points, the resulting charges have the same behavior as those obtained via vdW exclusion.

The final option explored here is to consider only the grid points corresponding to the CH_4 accessible volume in ZIF-8. These points are identified through Grand Canonical Monte Carlo (GCMC) simulations,[101] using a generic force field.[111,130] This approach simply gives fitted charges close to those obtained over the γ = 1.7 region. The same results are obtained introducing in the fit a weighting function for each grid point, according to the CH_4 probability density computed through GCMC simulations.

4. CONCLUSIONS

The problem of generating reliable fixed partial charges for periodic systems has been successfully tackled by introducing two major improvements to the well-established REPEAT method. The first improvement, D-REPEAT, consists of the simultaneous fit of the electrostatic potential (ESP), together with the total dipole fluctuations (TDF) of the crystal. The second, M-REPEAT, relies on the simultaneous fit of several ESP configurations at once. These two improvements can be fused together into the DM-REPEAT method, giving an effective tool to derive robust partial charges.

The statistical underdetermination of some atoms (i.e., buried atoms) in ZIF-8 has been investigated. This is particularly severe in the case of unrestrained REPEAT fitting performed over the crystallographic configuration, as first shown by Watanabe and co-workers,[51] where the resulting partial charges widely fluctuate as a function of γ (i.e., the vdW scaling factor controlling the region excluded from the fit). On one hand, the stability can be sensibly improved for the crystallographic configuration by coupling the corresponding ESP with the TDF. Still, some noise is present and the results are slightly biased toward the values of a pure TDF fit. On the other hand, stable results can also be found with a simultaneous fit of 100 ESP frames. In this case, the reproduction of the TDF is not optimal, with partial charges having a very small magnitude. The optimum, then, is found with the DM-REPEAT method, giving charges that are both stable over a large set of γ values and able to properly reproduce the total dipole fluctuations without spoiling the ESP. The same approach has been successfully applied to IRMOF-1 and ITQ-29, confirming the effectiveness of the proposed DM-REPEAT method.

The approach here presented is considered to be part of a complete force field parametrization (i.e., consisting of both bonded and nonbonded interactions). In this case, the collection of a large amount of QM data becomes cost-effective. Nonetheless, it can also be thought as a standalone task, where the reference QM data are generated over a small number of configurations taken from classical MD (using a raw force field), avoiding the expensive computation of a full BOMD trajectory.

Based on the findings of this work, and given the intrinsic difficulties of charges fitting, some key points should be taken

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

into account when performing this task. In fact, the fit of ESP data coming from the sole crystallographic configuration may lead to unphysical results. This can be alleviated by using a set of multiple (statistically independent) configurations. Moreover, it is of uttermost importance to enforce, whenever possible, the atom-type symmetry. Finally, it is more likely to obtain physically sound values when coupling together data coming from different but complementary properties, such as ESP and TDF.

■ ASSOCIATED CONTENT

Supporting Information

RRMS evolution plots in IRMOF-1 and ITQ-29. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00503.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: msant@uniss.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support by Italian Ministero dell'Istruzione, dell'Università e della Ricerca, by Università degli Studi di Sassari, by Istituto Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), and by Fondazione Banco di Sardegna is gratefully acknowledged. Cybersar Project managed by the Consorzio COSMOLAB is acknowledged for CPU time allocation. We thank Dr. M. Masia for useful discussions.

■ REFERENCES

(1) Breck, D. W. *Zeolite Molecular Sieves: Structure, Chemistry, and Use*; John Wiley & Sons: New York, 1973.
(2) Kärger, J.; Ruthven, D. M. *Diffusion in Zeolites and Other Microporous Solids*; John Wiley & Sons: New York, 1992.
(3) van Bekkum, H.et al., Eds. *Introduction to Zeolite Science and Practice*; Elsevier: Amsterdam, 2001.
(4) Rowsell, J. L.; Yaghi, O. M. *Microporous Mesoporous Mater.* **2004**, *73*, 3−14.
(5) Yang, Q.; Liu, D.; Zhong, C.; Li, J.-R. *Chem. Rev.* **2013**, *113*, 8261−8323.
(6) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. *Science* **2013**, *341*, 1230444-1−1230444-12.
(7) Rosi, N. L.; Eckert, J.; Eddaoudi, M.; Vodak, D. T.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. *Science* **2003**, *300*, 1127−1129.
(8) Rowsell, J. L. C.; Yaghi, O. M. *Angew. Chem., Int. Ed.* **2005**, *44*, 4670−4679.
(9) Bordiga, S.; Vitillo, J. G.; Ricchiardi, G.; Regli, L.; Cocina, D.; Zecchina, A.; Arstad, B.; Bjørgen, M.; Hafizovic, J.; Lillerud, K. P. *J. Phys. Chem. B* **2005**, *109*, 18237−18242.
(10) Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D.-W. *Chem. Rev.* **2012**, *112*, 782−835.
(11) Tranchemontagne, D. J.; Park, K. S.; Furukawa, H.; Eckert, J.; Knobler, C. B.; Yaghi, O. M. *J. Phys. Chem. C* **2012**, *116*, 13143−13151.
(12) Ren, J.; Langmi, H. W.; North, B. C.; Mathe, M. *Int. J. Energy Res.* **2015**, *39*, 607−620.
(13) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. *Energy Environ. Sci.* **2011**, *4*, 2177−2185.
(14) Li, J.-R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H.-K.; Balbuena, P. B.; Zhou, H.-C. *Coord. Chem. Rev.* **2011**, *255*, 1791−1823.
(15) Liu, Y.; Wang, Z. U.; Zhou, H.-C. *Greenhouse Gases: Sci. Technol.* **2012**, *2*, 239−259.
(16) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, *112*, 724−781.
(17) Huxford, R. C.; Della Rocca, J.; Lin, W. *Curr. Opin. Chem. Biol.* **2010**, *14*, 262−268.
(18) Keskin, S.; Kızılel, S. *Ind. Eng. Chem. Res.* **2011**, *50*, 1799−1812.
(19) Horcajada, P.; Gref, R.; Baati, T.; Allan, P. K.; Maurin, G.; Couvreur, P.; Férey, G.; Morris, R. E.; Serre, C. *Chem. Rev.* **2012**, *112*, 1232−1268.
(20) Bernini, M. C.; Fairen-Jimenez, D.; Pasinetti, M.; Ramirez-Pastor, A. J.; Snurr, R. Q. *J. Mater. Chem. B* **2014**, *2*, 766−774.
(21) Cunha, D.; Ben Yahia, M.; Hall, S.; Miller, S. R.; Chevreau, H.; Elkaïm, E.; Maurin, G.; Horcajada, P.; Serre, C. *Chem. Mater.* **2013**, *25*, 2767−2776.
(22) Koukaras, E. N.; Montagnon, T.; Trikalitis, P.; Bikiaris, D.; Zdetsis, A. D.; Froudakis, G. E. *J. Phys. Chem. C* **2014**, *118*, 8885−8890.
(23) He, L.; Wang, T.; An, J.; Li, X.; Zhang, L.; Li, L.; Li, G.; Wu, X.; Su, Z.; Wang, C. *CrystEngComm* **2014**, *16*, 3259−3263.
(24) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, 1989; pp 71−109, 126−131, 28−29, 159.
(25) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd Edition; Academic Press: Orlando, FL, USA, 2001; pp 63−105, 126−135 292−306.
(26) Snurr, R. Q.; Bell, A. T.; Theodorou, D. N. *J. Phys. Chem.* **1993**, *97*, 13742−13752.
(27) Demontis, P.; Suffritti, G. B. *Chem. Rev.* **1997**, *97*, 2845−2878.
(28) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(29) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(30) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(31) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205−1218.
(32) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(33) MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization, *Encyclopedia of Computational Chemistry*; John Wiley & Sons, Ltd.: Chichester, U.K., 2002; pp 271−277.
(34) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(35) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(36) Hirshfeld, F. *Theoret. Chim. Acta* **1977**, *44*, 129−138.
(37) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592−601.
(38) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304−323.
(39) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.
(40) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905.
(41) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(42) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.
(43) Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(44) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(45) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.
(46) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377−395.
(47) Sagui, C.; Pomorski, P.; Darden, T. A.; Roland, C. *J. Chem. Phys.* **2004**, *120*, 4530−4544.
(48) Tsiper, E. V.; Burke, K. *J. Chem. Phys.* **2004**, *120*, 1153−1156.
(49) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 1004−1013.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

(50) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455−2468.
(51) Watanabe, T.; Manz, T. A.; Sholl, D. S. *J. Phys. Chem. C* **2011**, *115*, 4824−4836.
(52) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2012**, *8*, 2844−2867.
(53) Xu, Q.; Zhong, C. *J. Phys. Chem. C* **2010**, *114*, 5035−5042.
(54) Zhang, P.; Bao, P.; Gao, J. *J. Comput. Chem.* **2011**, *32*, 2127−2139.
(55) Bultinck, P.; van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111-1−144111-9.
(56) Vanpoucke, D. E. P.; Bultinck, P.; van Driessche, I. *J. Comput. Chem.* **2013**, *34*, 405−417.
(57) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(58) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 1989−1998.
(59) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Theory Comput.* **2013**, *9*, 2221−2225.
(60) Jakobsen, S.; Jensen, F. *J. Chem. Theory Comput.* **2014**, *10*, 5493−5504.
(61) Hamad, S.; Balestra, S. R.; Bueno-Perez, R.; Calero, S.; Ruiz-Salvador, A. R. *J. Solid State Chem.* **2015**, *223*, 144−151.
(62) Wells, B. A.; De Bruin-Dickason, C.; Chaffee, A. L. *J. Phys. Chem. C* **2015**, *119*, 456−466.
(63) Ivanov, M. V.; Talipov, M. R.; Timerghazin, Q. K. *J. Phys. Chem. A* **2015**, *119*, 1422−1434.
(64) Kramer, C.; Spinn, A.; Liedl, K. R. *J. Chem. Theory Comput.* **2014**, *10*, 4488−4496.
(65) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866−2878.
(66) Chen, D.-L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225−10233.
(67) Thole, B.; van Duijnen, P. Th. *Theor. Chim. Acta* **1983**, *63*, 209−221.
(68) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893−928.
(69) Lillestolen, T. C.; Wheatley, R. J. *Chem. Commun.* **2008**, 5909−5911.
(70) Watanabe, T.; Sholl, D. S. *Langmuir* **2012**, *28*, 14114−14128.
(71) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, *4*, 3056−3061.
(72) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. *J. Phys. Chem. Lett.* **2012**, *3*, 2506−2511.
(73) Wilmer, C. E.; Snurr, R. Q. *Chem. Eng. J.* **2011**, *171*, 775−781.
(74) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 4313−4323.
(75) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253−287.
(76) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London, Ser. A* **1980**, *373*, 27−56.
(77) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345−8348.
(78) Heij, C.; de Boer, P.; Franses, P. H.; Kloek, T.; van Dijk, H. K. Multiple Regression. In *Econometric Methods with Applications in Business and Economics*; Oxford University Press: Oxford, U.K., 2004; pp 118−122.
(79) Spaldin, N. A. *J. Solid State Chem.* **2012**, *195*, 2−10.
(80) Resta, R.; Vanderbilt, D. *Physics of Ferroelectrics*; Springer: Berlin, 2007; pp 31−68.
(81) Berry, M. V. *Proc. R. Soc. London, Ser. A* **1984**, *392*, 45−57.
(82) Kudin, K. N.; Car, R.; Resta, R. *J. Chem. Phys.* **2007**, *126*, 234101-1−234101-8.
(83) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15−25.
(84) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(85) Frigo, M.; Johnson, S. *Proc. IEEE* **2005**, *93*, 216−231.
(86) Kolafa, J. *J. Comput. Chem.* **2004**, *25*, 335−342.
(87) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2003**, *118*, 4365−4369.
(88) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.
(89) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(90) Lippert, G.; Parrinello, M.; Hutter, J. *Mol. Phys.* **1997**, *92*, 477−488.
(91) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(92) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101-1−014101-7.
(93) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(94) Krack, M. *Theor. Chem. Acc.* **2005**, *114*, 145−152.
(95) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 3641−3662.
(96) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(97) Kosa, M.; Tan, J.-C.; Merrill, C. A.; Krack, M.; Cheetham, A. K.; Parrinello, M. *ChemPhysChem* **2010**, *11*, 2332−2336.
(98) Rana, M. K.; Pazzona, F. G.; Suffritti, G. B.; Demontis, P.; Masia, M. *J. Chem. Theory Comput.* **2011**, *7*, 1575−1582.
(99) Gabrieli, A.; Sant, M.; Demontis, P.; Suffritti, G. B. *Microporous Mesoporous Mater.* **2014**, *197*, 339−347.
(100) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104-1−154104-19.
(101) Gupta, A.; Chempath, S.; Sanborn, M. J.; Clark, L. A.; Snurr, R. Q. *Mol. Simul.* **2003**, *29*, 29−46.
(102) van Rossum, G. *Python Tutorial*; Report CS-R9526, Centrum voor Wiskunde en Informatica: Amsterdam, 1995.
(103) *The Python Language Reference*. Available via the Internet at http://docs.python.org/release/2.7 (accessed July 7, 2015).
(104) Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10186−10191.
(105) Banerjee, R.; Phan, A.; Wang, B.; Knobler, C.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. *Science* **2008**, *319*, 939−943.
(106) Phan, A.; Doonan, C. J.; Uribe-Romo, F. J.; Knobler, C. B.; O'Keeffe, M.; Yaghi, O. M. *Acc. Chem. Res.* **2010**, *43*, 58−67.
(107) Moggach, S.; Bennett, T.; Cheetham, A. *Angew. Chem., Int. Ed.* **2009**, *48*, 7087−7089.
(108) Pérez-Pellitero, J.; Amrouche, H.; Siperstein, F.; Pirngruber, G.; Nieto-Draghi, C.; Chaplais, G.; Simon-Masseron, A.; Bazer-Bachi, D.; Peralta, D.; Bats, N. *Chem.—Eur. J.* **2010**, *16*, 1560−1571.
(109) Hu, Y.; Kazemian, H.; Rohani, S.; Huang, Y.; Song, Y. *Chem. Commun.* **2011**, *47*, 12694−12696.
(110) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Düren, T. *J. Am. Chem. Soc.* **2011**, *133*, 8900−8902.
(111) Zheng, B.; Sant, M.; Demontis, P.; Suffritti, G. B. *J. Phys. Chem. C* **2012**, *116*, 933−938.
(112) Pusch, A.-K.; Splith, T.; Moschkowitz, L.; Karmakar, S.; Biniwale, R.; Sant, M.; Suffritti, G.; Demontis, P.; Cravillon, J.; Pantatosaki, E.; Stallmach, F. *Adsorption* **2012**, *18*, 359−366.
(113) Zhang, L.; Hu, Z.; Jiang, J. *J. Am. Chem. Soc.* **2013**, *135*, 3722−3728.
(114) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F.-X. *J. Phys. Chem. Lett.* **2013**, *4*, 1861−1865.
(115) Pantatosaki, E.; Jobic, H.; Kolokolov, D. I.; Karmakar, S.; Biniwale, R.; Papadopoulos, G. K. *J. Chem. Phys.* **2013**, *138*, 034706-1−034706-10.
(116) Zheng, B.; Pan, Y.; Lai, Z.; Huang, K.-W. *Langmuir* **2013**, *29*, 8865−8872.
(117) Tanaka, H.; Ohsaki, S.; Hiraide, S.; Yamamoto, D.; Watanabe, S.; Miyahara, M. T. *J. Phys. Chem. C* **2014**, *118*, 8445−8454.
(118) Demontis, P.; Suffritti, G. B. *Microporous Mesoporous Mater.* **2009**, *125*, 160−168.
(119) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402*, 276−279.
(120) Dixit, M.; Adit Maark, T.; Ghatak, K.; Ahuja, R.; Pal, S. *J. Phys. Chem. C* **2012**, *116*, 17336−17342.
(121) Corma, A.; Rey, F.; Rius, J.; Sabater, M. J.; Valencia, S. *Nature* **2004**, *431*, 287−290.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843

(122) Baerlocher, C.; McCusker, L.; Olson, D. *Atlas of Zeolite Framework Types*; Elsevier Science: Amsterdam, 2007; pp 194−195.
(123) Tiscornia, I.; Valencia, S.; Corma, A.; Téllez, C.; Coronas, J.; Santamaría, J. *Microporous Mesoporous Mater.* **2008**, *110*, 303−309.
(124) Huang, A.; Caro, J. *Chem. Commun.* **2010**, *46*, 7748−7750.
(125) García-Sánchez, A.; Ania, C. O.; Parra, J. B.; Dubbeldam, D.; Vlugt, T. J. H.; Krishna, R.; Calero, S. *J. Phys. Chem. C* **2009**, *113*, 8814−8820.
(126) Gabrieli, A.; Sant, M.; Demontis, P.; Suffritti, G. B. *J. Phys. Chem. C* **2013**, *117*, 503−509.
(127) Ghermani, N. E.; Lecomte, C.; Dusausoy, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *53*, 5231−5239.
(128) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(129) Doemer, M.; Maurer, P.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U. *J. Chem. Theory Comput.* **2014**, *10*, 412−422.
(130) Goodbody, S. J.; Watanabe, K.; MacGowan, D.; Walton, J. P. R. B.; Quirke, N. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 1951−1958.

DOI: 10.1021/acs.jctc.5b00503
J. Chem. Theory Comput. 2015, 11, 3829−3843