

# Expanding the Scope of Density Derived Electrostatic and Chemical Charge Partitioning to Thousands of Atoms

Louis P. Lee,<sup>†</sup> Nidia Gabaldon Limas,<sup>‡</sup> Daniel J. Cole,<sup>\*,§</sup> Mike C. Payne,<sup>†</sup> Chris-Kriton Skylaris,<sup>||</sup> and Thomas A. Manz<sup>\*,‡</sup>

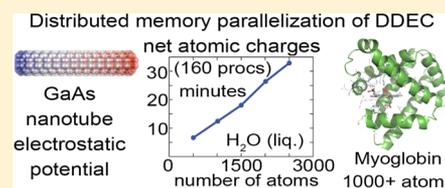
<sup>†</sup>TCM Group, Cavendish Laboratory, 19 JJ Thomson Ave, Cambridge CB3 0HE, United Kingdom

<sup>‡</sup>Chemical & Materials Engineering Department, New Mexico State University, Las Cruces, New Mexico 88003-8001, United States

<sup>§</sup>Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States

<sup>||</sup>School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

**ABSTRACT:** The density derived electrostatic and chemical (DDEC/c3) method is implemented into the ONETEP program to compute net atomic charges (NACs), as well as higher-order atomic multipole moments, of molecules, dense solids, nanoclusters, liquids, and biomolecules using linear-scaling density functional theory (DFT) in a distributed memory parallel computing environment. For a >1000 atom model of the oxygenated myoglobin protein, the DDEC/c3 net charge of the adsorbed oxygen molecule is approximately  $-1e$  (in agreement with the Weiss model) using a dynamical mean field theory treatment of the iron atom, but much smaller in magnitude when using the generalized gradient approximation. For GaAs semiconducting nanorods, the system dipole moment using the DDEC/c3 NACs is about 5% higher in magnitude than the dipole computed directly from the quantum mechanical electron density distribution, and the DDEC/c3 NACs reproduce the electrostatic potential to within approximately 0.1 V on the nanorod's solvent-accessible surface. As examples of conducting materials, we study (i) a 55-atom Pt cluster with an adsorbed CO molecule and (ii) the dense solids Mo<sub>2</sub>C and Pd<sub>3</sub>V. Our results for solid Mo<sub>2</sub>C and Pd<sub>3</sub>V confirm the necessity of a constraint enforcing exponentially decaying electron density in the tails of buried atoms.



## 1. INTRODUCTION

Atoms-in-molecule (AIM) partitioning of the quantum mechanical electron density is important both for understanding the chemical nature of materials and for developing force fields for use in classical molecular dynamics and Monte Carlo simulations.<sup>1</sup> Because there is some flexibility in how to partition electron density among the atoms in a material, the key criterion is to maximize usefulness. In this regard, the AIM densities should: (a) produce net atomic charges (NACs) that are chemically meaningful by being compatible with electronegativity scales, (b) yield an efficiently converging atom-centered multipole expansion of the electrostatic potential surrounding the material, which is important for constructing force fields for atomistic simulations, (c) be relatively insensitive to different thermodynamically accessible molecular conformations, which is needed to construct transferable flexible force fields, (d) be inexpensive to compute, and (e) be applicable to a wide range of materials (such as molecules, solids, and so on). Regarding point c, the electron distribution is dependent on the molecule's environment and geometry, and thus polarizable force fields may be required to further improve transferability of the charge model or multipolar expansion.<sup>2</sup> Once an atom's electron density distribution  $n_A(\mathbf{r})$  in a material has been found, the NAC of atom A,  $Q_A$ , is then obtained as

$$Q_A = z_A - N_A \quad (1)$$

where  $N_A$  is the number of electrons assigned to atom A, and  $z_A$  is its effective nuclear charge.

At the most basic level, one can ask whether atoms should be overlapping or nonoverlapping. The answer to this question depends in part on whether the NACs are intended for use in constructing force fields for classical atomistic simulations. By Gauss' Law of Electrostatics, the NACs will most accurately reproduce  $V(\mathbf{r})$  surrounding the material if  $\{n_A(\mathbf{r})\}$  is close to spherically symmetric, and this criterion can only be fulfilled using overlapping atoms. Specifically, Gauss' Law of Electrostatics states that the integral of the electric field over a closed surface is proportional to the enclosed charge. Therefore, by symmetry an enclosed spherically symmetric charge density distribution yields the same electrostatic potential outside the bounding surface as an equivalent point charge placed at its center. Thus, when assigning nearly spherical atomic electron density distributions, the electrostatic potential surrounding the material can be approximated by an atom-centered point charge model, and the atomic multipole moments will be small.

Without any loss of generality,  $n_A(\mathbf{r})$  can be defined in terms of a weighting factor,  $w_A(\mathbf{r}) \geq 0$ , such that

$$n_A(\mathbf{r}) = \frac{w_A(\mathbf{r})}{\sum_{k,B} w_B(\mathbf{r})} n(\mathbf{r}) \quad (2)$$

Received: August 25, 2014

Published: November 3, 2014

where  $n(\mathbf{r})$  is the system's electron density at a position  $\mathbf{r}$ , and the sum in the denominator of eq 2 runs over all atoms in the unit cell plus their images along any directions that are periodic. Note that eq 2 guarantees that the  $\{n_A(\mathbf{r})\}$  sum to  $n(\mathbf{r})$  at every position  $\mathbf{r}$ . Thus, the task for defining  $\{n_A(\mathbf{r})\}$  has been transformed into the task of defining  $\{w_A(\mathbf{r})\}$ . Various definitions for  $\{w_A(\mathbf{r})\}$  have been proposed, leading to different AIM methods. However, one feature that AIM methods have in common is that they yield global minima of the optimization functional  $\Omega$  and stationary points of the path action  $S$  proposed by Manz:

$$\Omega = \sum_A \oint P(\xi_A(\mathbf{r})) d^3\mathbf{r} \quad (3)$$

$$\xi_A(\mathbf{r}) = \frac{n_A(\mathbf{r}) \sum_{k,B} w_B(\mathbf{r})}{w_A(\mathbf{r}) n(\mathbf{r})} \quad (4)$$

$$P(\xi) = \xi \ln(\xi) - \xi + 1 \quad (5)$$

$$dS = \sum_A \oint \ln(\xi_A(\mathbf{r})) \delta n_A(\mathbf{r}) d^3\mathbf{r} \quad (6)$$

where the action  $S = \int dS$ , takes a path in  $\{n_A(\mathbf{r})\}$  optimization space as its argument and has a real number as its result.<sup>3</sup>

For AIM methods with nonoverlapping atoms, such as Bader's quantum chemical topology,  $w_A(\mathbf{r}) \in \{0,1\}$ .<sup>4</sup> For AIM methods with overlapping atoms,  $w_A(\mathbf{r})$  is chosen to be a continuous function. A key observation is that if  $\sum_{k,B} w_B(\mathbf{r})$  is made to resemble  $n(\mathbf{r})$ , then  $n_A(\mathbf{r})$  will resemble  $w_A(\mathbf{r})$ . As described above,  $\{n_A(\mathbf{r})\}$  should be made close to spherically symmetric to yield a rapidly converging atom-centered multipole expansion that reproduces  $V(\mathbf{r})$  surrounding the material. Hence, it is reasonable to define  $\{w_A(\mathbf{r})\}$  as a set of spherically symmetric functions,  $\{w_A(r_A)\}$ , where  $r_A = |\mathbf{r} - \mathbf{R}_A|$  and  $\mathbf{R}_A$  is the position of the nucleus of atom  $A$ .

In section 2.1, we summarize the advantages and limitations of various definitions of  $\{w_A(r_A)\}$ . In particular, we describe the basic principles of *density derived electrostatic and chemical* (DDEC) atomic population analysis. The main idea of the DDEC method is to simultaneously optimize each weighting factor  $w_A(r_A)$  to resemble (a) the spherical average of  $n_A(\mathbf{r})$  in order to yield a rapidly converging atom-centered multipole expansion that reproduces  $V(\mathbf{r})$  surrounding the material and (b) the density of a reference ion of the same element having the same number of electrons,  $N_A$ , in order to make the assigned atomic charges chemically meaningful with good conformational transferability.

The DDEC method has been previously used to construct force fields for proteins,<sup>5</sup> metal–organic frameworks,<sup>1,6–12</sup> zeolites,<sup>13–15</sup> and ionic liquids<sup>16</sup> and to study catalysts and other advanced materials.<sup>17–25</sup> The method has gone through several generations of improvements, in which the third generation (named DDEC/c3) combines and extends some features of earlier methodologies to produce accurate results across a wider range of material types.<sup>3,26</sup> In a previous article,<sup>5</sup> some of us programmed a second generation variant, which we here name the DDEC/cc2 method, into the ONETEP density functional theory (DFT) program.<sup>27</sup> In this article, we implement the DDEC/c3 method into the ONETEP program to achieve the following improvements relative to the DDEC/cc2 method: (a) all-electron density partitioning, (b) use of an improved precomputed reference density library, and (c)

enforcement of exponential decay constraints to prevent buried atoms from becoming too diffuse. These improvements are especially important for dense solids containing short bond lengths, such as the Mo<sub>2</sub>C and Pd<sub>3</sub>V solids that are studied in section 4.

ONETEP is a linear-scaling DFT program that uses a distributed memory parallel computing environment.<sup>28,29</sup> ONETEP combines high basis set accuracy, comparable to that of plane-wave DFT methods, with a computational cost that scales linearly with the number of atoms in the system, thus allowing accurate DFT calculations to be performed on systems comprising many thousands of atoms.<sup>30–34</sup> By implementing the DDEC/c3 method into ONETEP, we take advantage of ONETEP's distributed memory model to achieve parallelization over multiple compute nodes and a large number of processors. This facilitates rapid charge analysis on a much larger scale than was previously possible and allows us to perform chemical analysis and, potentially, force field design for biomolecules, nanoclusters, liquids, and other interesting materials.

Although most of our analysis in this paper will concentrate on the NACs that are derived from the DDEC/c3 electron density partitioning, we emphasize that there are many well-documented cases where atom-centered point charges inadequately describe the complexity of the quantum mechanical  $V(\mathbf{r})$ .<sup>35</sup> For example, in the hollow single-walled boron-nitride nanotube studied in refs 26 and 3, the electrostatic potential inside the nanotube is approximately 0.6 volts higher than that outside the nanotube. Because an atom-centered point charge model necessarily yields equal electrostatic potentials inside and outside the nanotube, the electrostatic potential of such a system cannot be described by any atom-centered point charge model. For this reason, the DDEC method uses an atom-centered multipole expansion to represent the electrostatic potential. For the boron-nitride nanotube, the root-mean-square error (RMSE) in the electrostatic potential dropped from 8.81 kcal/mol (DDEC/c3 NACs only) to 2.40 kcal/mol (DDEC/c3 NACs plus atomic dipoles).<sup>3</sup> This clearly illustrates the importance of including atomic multipoles when constructing force-fields for some materials, and in section 5.2 of this paper we will describe a further case where atomic dipoles cannot be neglected.

The remainder of this article is organized as follows. Section 2.1 summarizes the DDEC/c3 method and its relationship to earlier AIM charge partitioning schemes. Section 2.2 describes implementation of the DDEC/c3 method into the ONETEP program. Section 3 summarizes computational parameters. Section 4 validates the computational approach through comparisons to various benchmark calculations. Section 5 applies the charge analysis method to typical problems in nanomaterials and biology, namely, a 55-atom Pt cluster with an adsorbed CO molecule, GaAs polar nanorods, and an oxygenated myoglobin molecule comprising more than 1000 atoms.

## 2. THEORY

**2.1. Density Derived Electrostatic and Chemical Charges.** The DDEC method<sup>3,26</sup> is a density-based AIM charge partitioning scheme that combines the *iterative Hirshfeld*<sup>36</sup> (IH) and *iterated stockholder atoms*<sup>37</sup> (ISA) methods. Both IH and ISA schemes in turn originate from the Hirshfeld charge partitioning method.<sup>38</sup> In the original Hirshfeld formulation, for each atom  $A$ , the electron density  $n(\mathbf{r})$  is divided into overlapping atomic densities  $n_A(\mathbf{r})$ :

$$n_A(\mathbf{r}) = \frac{n_A^0(r_A)}{\sum_{k,B} n_B^0(r_B)} n(\mathbf{r}) \quad (7)$$

where  $n_A^0(r_A)$  is a spherically symmetric reference atomic density, whose sum over all atoms in the unit cell, plus their images along any periodic directions, is termed the promolecular density. Nalewajski and Parr showed that this form of stockholder partitioning minimizes the distance  $F^{\text{AIM}}$  between the real and promolecular density in an information-theoretic sense:<sup>39</sup>

$$F^{\text{AIM}} = \sum_A \oint n_A(\mathbf{r}) \ln \left[ \frac{n_A(\mathbf{r})}{n_A^0(r_A)} \right] d^3\mathbf{r} \quad (8)$$

where the sum is performed over all atoms A in the unit cell and  $\oint d^3\mathbf{r}$  denotes volume integration over all space. It has since been shown that the Hirshfeld partitioning can also be obtained via minimization of the Hellinger–Bhattacharya distance metric.<sup>40</sup> In the original Hirshfeld method, neutral gas-phase atomic densities were typically chosen as the atomic weighting factors, leading to assigned atomic populations that were often too close to zero.<sup>3,26,36,41</sup> To address this problem, Bultinck et al.<sup>36</sup> proposed an iterative Hirshfeld method in which the partitioned atomic densities for each new iteration,  $n_A^{i+1}(\mathbf{r})$ , are computed as

$$n_A^{i+1}(\mathbf{r}) = \frac{w_A^i(r_A)}{\sum_{k,B} w_B^i(r_B)} n(\mathbf{r}) \quad (9)$$

where  $w_A^i(r_A)$  is the atomic weighting factor that is calculated from the partitioned atomic density  $n_A^i(\mathbf{r})$  of the previous iteration  $i$ . In the IH scheme, the spherically symmetric weighting factors  $w_A^i(r_A) = w_A^{\text{IH},i}(r_A)$  are constructed via linear interpolation between reference densities  $n_A^0(\tau, r_A)$  of free atoms or ions with the next lowest integer ( $\tau = \text{lint}(N_A^i)$ ) and the next highest integer ( $\tau + 1$ ) number of electrons:

$$w_A^{\text{IH},i}(r_A) = n_A^0(\tau, r_A) [\tau + 1 - N_A^i] + n_A^0(\tau + 1, r_A) [N_A^i - \tau] \quad (10)$$

in order to obtain a free reference density for a hypothetical ion comprising  $N_A^i$  electrons. The procedure is iterated until the changes in the IH NACs converge below a specified threshold. The IH NACs have been shown to reproduce *ab initio* electrostatic properties of small molecules and to show good conformational transferability.<sup>3,42,43</sup> However, using isolated reference anions that are much more diffuse than in condensed materials has been shown to lead to chemically unreasonable NACs in certain situations.<sup>26,44,45</sup>

The iterated stockholder atoms (ISA) method<sup>37</sup> takes the spherical average of  $n_A^i(\mathbf{r})$  as the weighting factor  $w_A^i(r_A) = w_A^{\text{ISA},i}(r_A)$  that enters into eq 9:

$$w_A^{\text{ISA},i}(r_A) = \langle n_A^i(\mathbf{r}) \rangle_A = \oint \frac{\delta^{\text{dirac}}(r'_A - r_A)}{4\pi r_A^2} n_A^i(\mathbf{r}') d^3\mathbf{r}' \quad (11)$$

where  $\langle \dots \rangle_A$  denotes spherical averaging about the center of atom A. The averaging is performed on a set of discrete radial shells up to a predefined maximum radius. The ISA scheme produces a better fit to the *ab initio* electrostatic potential than the IH method,<sup>5,26</sup> which is due to the spherical-averaging procedure used to generate the atomic weighting factors. However, the ISA electron density distributions are not

constrained to decay like those of real atoms, which can lead to inaccurate and poorly transferable NACs in systems with buried atoms.<sup>3,5,26,46</sup>

In the original formulation of the DDEC scheme by Manz and Sholl,<sup>3,26</sup> the IH and ISA methods are mixed via minimization of a combined information entropy functional (eq 8):

$$F^{\text{DDEC}} = \chi F^{\text{IH}} + (1 - \chi) F^{\text{ISA}} \quad (12)$$

where  $F^{\text{IH/ISA}}$  are constructed with IH/ISA reference densities and  $\chi$  is an adjustable mixing parameter. Minimizing the partial derivative of eq 12 with respect to  $n_A(\mathbf{r})$  leads to the following partitioning:

$$n_A^{i+1}(\mathbf{r}) = \frac{w_A^{c2,i}(r_A)}{\sum_{k,B} w_B^{c2,i}(r_B)} n(\mathbf{r}) \quad (13)$$

$$w_A^{c2,i}(r_A) = [w_A^{\text{IH},i}(r_A)]^\chi [w_A^{\text{ISA},i}(r_A)]^{1-\chi} \quad (14)$$

where  $w_A^{\text{IH/ISA},i}(r_A)$  are the spherically symmetric weighting factors given by eqs 10 and 11, and the label c2 denotes the original DDEC/c2 scheme of Manz and Sholl.<sup>26</sup> The DDEC/c2 scheme allows a fraction of  $F^{\text{IH}}$  to contribute toward curvature in regions that otherwise have shallow optimization landscapes, for example for buried atoms, thus alleviating the slow convergence of the ISA method while retaining the low-order multipoles possessed by the converged  $\{n_A(\mathbf{r})\}$ .

In the original DDEC/c2 scheme,<sup>26</sup> the IH atomic reference densities  $n_A^0(\tau, r_A)$  (eq 10) are generated from DFT ground state calculations performed in the presence of a charge compensation sphere. The reason for using compensated densities, instead of free ionic states as in the original IH scheme, is to mimic the dielectric screening experienced by an ion embedded within a condensed material, whereby its density profile is modified by the effective dielectric constant of the material. The charge compensation sphere expands (contracts) the reference electron density of cations (anions) and also ensures stability of certain anionic species, such as  $\text{O}^{2-}$ , whose outer electrons would otherwise be unbound. This combined charge partitioning and reference density generation scheme, called DDEC/c2, has been shown to perform well for a variety of systems, ranging from periodic bulk crystals and slabs to isolated molecules, in terms of producing chemically meaningful atomic charges that reproduce the electrostatic potential of the system.<sup>26</sup>

Despite its successes, for certain compacted materials with short bond lengths and atoms with diffuse electron densities, the DDEC/c2 method has been observed to converge slowly to unrealistic NACs,<sup>3</sup> even with the additional IH weighting that is introduced to counter the shallow charge optimization landscape of the ISA method. For example, charge partitioning in crystalline  $\text{Pd}_3\text{V}$  yields an unrealistic charge of  $-0.98e$  on the more electropositive V, and runaway charges exceeding  $-6e$  were observed for carbon in crystalline  $\text{Mo}_2\text{C}$ .<sup>3</sup> These issues stem from the excessive overlap between diffuse atoms due to their short bond lengths and are addressed in the recently published DDEC/c3 scheme.<sup>3</sup> Here, we provide a short summary of the additional constraints that are incorporated into the DDEC/c3 method.

First, the c3 reference densities  $n_A^{c3,0}(\tau, r_A)$  are derived from the c2 reference densities by enforcing the following three constraints:

$$\frac{\partial n_A^{c3,0}(\tau, r_A)}{\partial r_A} \leq 0, \quad \frac{\partial n_A^{c3,0}(\tau, r_A)}{\partial N_A} \geq 0, \quad \frac{\partial^2 n_A^{c3,0}(\tau, r_A)}{\partial r_A \partial N_A} \leq 0 \quad (15)$$

In practice, all-electron c2 reference densities are already monotonically decaying with respect to  $r_A$ , leaving only two remaining constraints to be enforced for each  $r_A$ , which is performed as follows. The neutral reference density,  $n_A^{c3,0}(\tau = z_A, r_A)$ , is set equal to  $n_A^{c3,0}(\tau = z_A, r_A)$  and remains unchanged. Next, for the  $\pm 1$  ions, the reference densities  $n_A^{c3,0}(\tau = z_A \pm 1, r_A)$  are adjusted by subtracting density from the +1 and adding density to the -1 ion, such that, with respect to the neutral species:

$$\begin{aligned} & (n_A^{c3,0}(\tau + 1, r_A + \Delta r_A) - n_A^{c3,0}(\tau + 1, r_A)) \\ & - (n_A^{c3,0}(\tau, r_A + \Delta r_A) - n_A^{c3,0}(\tau, r_A)) \leq 0 \end{aligned} \quad (16)$$

The density of each adjusted atomic reference density is then normalized such that

$$\int_0^\infty 4\pi r_A^2 n_A^{c3,0}(\tau, r_A) dr_A = \tau \quad (17)$$

These two steps (eqs 16 and 17) are iterated until the density profiles  $n_A^{c3,0}(\tau = z_A \pm 1, r_A)$  converge. The entire procedure is then repeated for the  $\pm 2$  ionic species, starting from the converged  $n_A^{c3,0}(\tau = z_A \pm 1, r_A)$ , and continues until the density profiles of all relevant ionic species have been obtained.

Second, the IH weighting factor  $w_A^{IH,i}(r_A)$  that enters into eq 14 for every iteration is replaced by a *conditioned reference density*  $Y_A^{avg,i}(r_A)$ , which is computed as the spherical average of  $n_A^{IH,i}(\mathbf{r})$ :

$$\begin{aligned} Y_A^{avg,i}(r_A) &= \langle n_A^{IH,i}(\mathbf{r}) \rangle_A \\ &= w_A^{IH,i-1}(r_A) \left\langle \frac{n(\mathbf{r})}{\sum_{k,B} w_B^{IH,i-1}(r_B)} \right\rangle_A \end{aligned} \quad (18)$$

$$\sigma_A^i(r_A) = [Y_A^{avg,i}(r_A)]^\gamma [w_A^{ISA,i}(r_A)]^{1-\gamma} \quad (19)$$

where  $\gamma$  is the DDEC/c3 mixing parameter. This modification ensures that the sum of (initial) DDEC weighting factors  $\{\sigma_A^i(r_A)\}$  remains consistently close to the real density  $n(\mathbf{r})$ , as the density profile of  $Y_A^{avg,i}(r_A)$  is derived from the partitioning of  $n(\mathbf{r})$  itself and not merely an interpolation between two reference densities, and further enhances the curvature of the charge optimization landscape.<sup>3</sup> In addition, this conditioning process yields an optimum value of  $\gamma = 3/14$ , which is independent of the system being studied, and has been shown to result in a good balance between minimizing the atomic multipoles of the partitioned atomic densities and maximizing chemical accuracy.<sup>3,5</sup>

Third, the DDEC/c3 method incorporates additional constraints to enforce exponential decay of the atomic densities with increasing distance from each atomic center. These constraints have the greatest effect on the tail regions, which can become deeply buried in the partitioned atomic densities of neighboring atoms in nonporous systems:

$$\lim_{w_A^{c3}(r_A) \ll w^{c3}(\mathbf{r})} w_A^{c3}(r_A) \simeq w_A^{c3}(r_A^{\text{buried}}) e^{-(r_A - r_A^{\text{buried}})t_A} \quad (20)$$

where  $r_A > r_A^{\text{buried}}$  defines a buried region and  $t_A$  is an effective decay exponent, which is constrained to prevent a buried tail from becoming too diffuse. Meanwhile,  $w^{c3}(\mathbf{r}) = \sum_{k,B} w_B^{c3}(r_B)$ ,

with  $w_A^{c3}(r_A)$  being the new DDEC/c3 weighting factors used to partition the atomic density. Equation 20 implies a constraint of the form

$$\frac{\partial w_A^{c3}(r_A)}{\partial r_A} + \eta_A(r_A) w_A^{c3}(r_A) \leq 0 \quad (21)$$

In the limit of  $n_A(r_A) \rightarrow n(\mathbf{r})$ , however,  $w_A^{c3}(\mathbf{r})$  should only be constrained to decay monotonically, and the exponential constraint should not be applied. A suitable function  $\eta_A(r_A)$  that satisfies both of these conditions is given by

$$\eta_A(r_A) = b \{1 - [\zeta_A(r_A)]^2\} \quad (22)$$

$$\zeta_A(r_A) = Y_A^{avg}(r_A) \frac{\langle [Y^{avg}(\mathbf{r})]^{-1/2} \rangle_A}{\langle [Y^{avg}(\mathbf{r})]^{1/2} \rangle_A} \quad (23)$$

where  $Y^{avg}(\mathbf{r}) = \sum_{k,B} Y_B^{avg}(r_B)$ . The form of eq 22, with  $b = 1.75$  Bohr<sup>-1</sup>, has been shown to be optimal in ensuring good transferrability of the atomic charge distributions.<sup>3</sup>

In calculations using frozen core electrons, the total number of electrons assigned to a particular atom must be at least as large as the number of core electrons assigned to that atom:

$$N_A^{\text{val}} = N_A - N_A^{\text{core}} \geq 0 \quad (24)$$

where  $N_A^{\text{val}}$ ,  $N_A$ , and  $N_A^{\text{core}}$  are the number of valence, total, and core electrons assigned to atom A, respectively. Equation 24 is enforced by expressing  $w_A^{c3}(r_A)$  in terms of an auxiliary radial function  $G_A(r_A)$ :

$$w_A^{c3}(r_A) = \lambda_A G_A(r_A) \quad (25)$$

$$\lambda_A = \max \left\{ N_A^{\text{core}} \left[ \oint \frac{G_A(r_A)}{w^{c3}(\mathbf{r})} n(\mathbf{r}) d^3\mathbf{r} \right]^{-1}, 1 \right\} \quad (26)$$

$\lambda_A$  serves to rescale  $G_A(r_A)$  by a constant factor whenever  $\oint ((G_A(r_A))/(w^{c3}(\mathbf{r}))) n(\mathbf{r}) d^3\mathbf{r} < N_A^{\text{core}}$ , in order to prevent the total integrated partitioned atomic density  $N_A$  from being less than  $N_A^{\text{core}}$ .  $G_A(r_A)$  is constrained to be normalized with respect to  $\sigma_A(r_A)$ :

$$\int_0^\infty 4\pi r_A^2 [G_A(r_A) - \sigma_A(r_A)] dr_A = 0 \quad (27)$$

The function  $G_A(r_A)$  is obtained as a weighted least-squares minimization with respect to  $\sigma_A(r_A)$ , by extremizing the functional:<sup>3</sup>

$$\begin{aligned} H[G_A(r_A)] &= \frac{1}{2} \int_0^\infty 4\pi r_A^2 \frac{[G_A(r_A) - \sigma_A(r_A)]^2}{[\sigma_A(r_A)]^{1/2}} dr_A \\ &+ \int_0^\infty 4\pi r_A^2 \Gamma_A(r_A) \left\{ \frac{dG_A(r_A)}{dr_A} + \eta_A(r_A) G_A(r_A) \right\} dr_A \\ &- \Phi_A \int_0^\infty 4\pi r_A^2 [G_A(r_A) - \sigma_A(r_A)] dr_A \end{aligned} \quad (28)$$

The second and third terms of eq 28 enforce the constraints in eqs 21 and 27, respectively, with  $\Gamma_A(r_A)$  and  $\Phi_A$  being their associated Lagrange multipliers. Minimization of eq 28 leads to an expression for  $G_A(r_A)$  in terms of  $\sigma_A(r_A)$ :

$$G_A(r_A) = \sigma_A(r_A) + [\sigma_A(r_A)]^{1/2} \left\{ \Phi_A - \Gamma_A(r_A) \eta_A(r_A) + \frac{d\Gamma_A(r_A)}{dr_A} + \frac{2\Gamma_A(r_A)}{r_A} \right\} \quad (29)$$

In practice, an integration strategy is required to implement the above equations. Possible integration strategies include overlapping atom-centered grids,<sup>44,47</sup> nonoverlapping atom-centered grids,<sup>48</sup> and uniform grids with valence-core separation.<sup>3</sup> The integration strategy affects the computational cost and precision without changing the underlying definition of the charge partitioning method. In this work, we use uniform grids with valence-core separation, because the ONETEP program already generates and uses the valence electron density on a uniformly spaced Cartesian grid. However, the cusp in  $n(\mathbf{r})$  close to the nucleus of heavy atoms in all-electron calculations makes integration of the core density inaccurate, unless an extremely fine grid is used. For an all-electron calculation, the partitioned atomic densities are therefore separated into valence and core contributions:  $n_A^{\text{val}}(\mathbf{r}) = n_A(\mathbf{r}) - n_A^{\text{core}}(\mathbf{r})$ .<sup>26</sup> The ISA scheme is used to partition the core density  $n_A^{\text{core}}(\mathbf{r}) = n(\mathbf{r}) - n_A^{\text{val}}(\mathbf{r})$  into  $n_A^{\text{core}}(r_A)$ , while the DDEC scheme is used to partition the total density  $n(\mathbf{r})$ . Note that the constraint in eq 21 is also applied to the updated core ISA weighting factor  $w_A^{\text{ISA,core}}(r_A)$  at every iteration, with  $\eta_A(r_A) = b_{\text{core}} = 2.0 \text{ Bohr}^{-1}$ .<sup>49</sup> NACs are then obtained by inserting  $N_A = N_A^{\text{val}} + N_A^{\text{core}}$  into eq 1, where  $N_A^{\text{val}}$  is the volume integral of  $n_A^{\text{val}}(\mathbf{r})$  and  $N_A^{\text{core}}$  is the analytic value based on chemical knowledge and the nature of the initial valence-core separation of  $n(\mathbf{r})$ . This ensures that the potentially inaccurate representation of  $n_A^{\text{core}}(\mathbf{r})$  and  $n_A^{\text{core}}(r_A)$  does not affect the accuracy of the partitioned NACs.

**2.2. DDEC/c3 Implementation in ONETEP.** ONETEP<sup>27</sup> is a linear-scaling DFT package which is based on a reformulation of conventional Kohn–Sham DFT:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha, \beta} \varphi_{\alpha}(\mathbf{r}) K^{\alpha\beta} \varphi_{\beta}^*(\mathbf{r}') \quad (30)$$

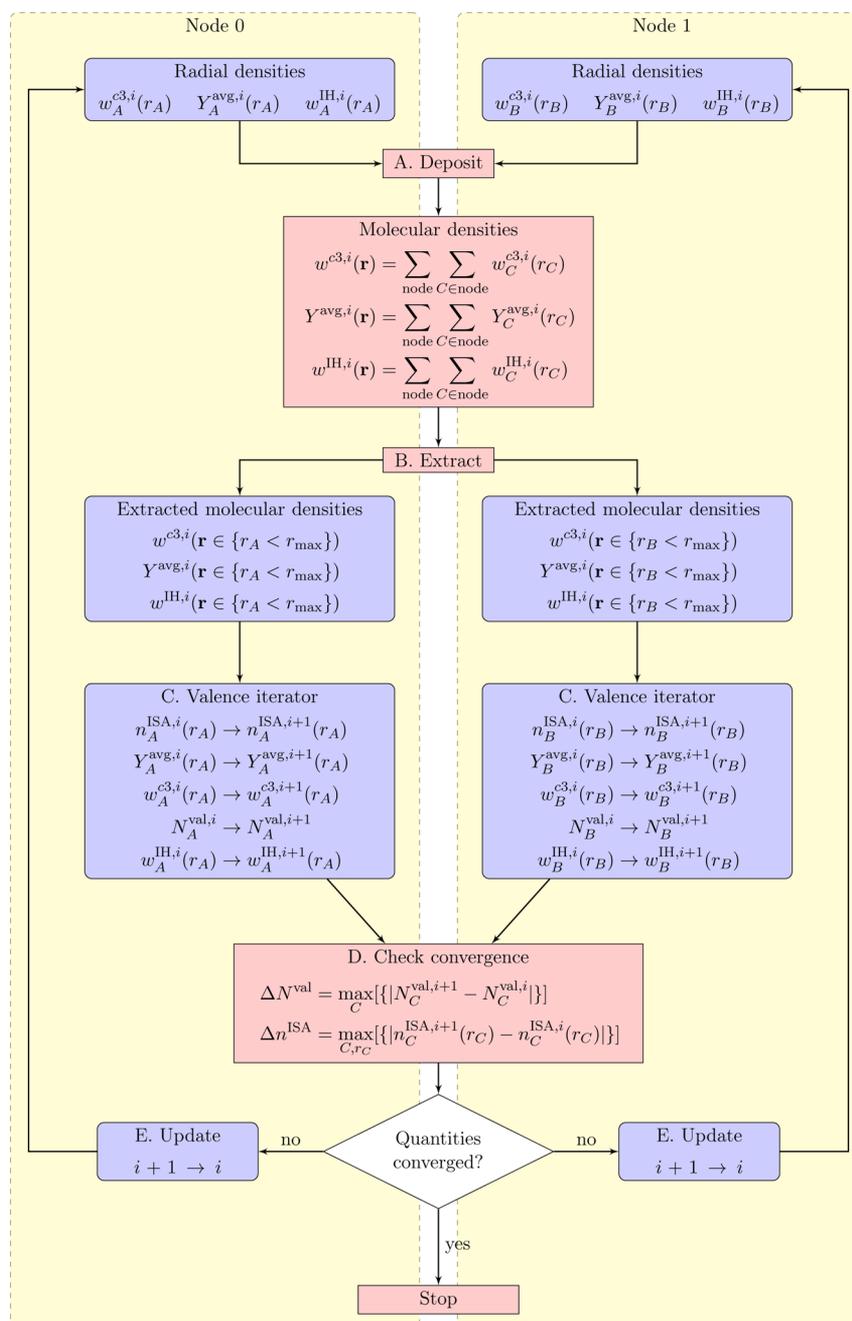
where  $\{\varphi_{\alpha}(\mathbf{r})\}$  are a minimal set of spatially localized nonorthogonal generalized Wannier functions (NGWFs),<sup>50</sup>  $\rho(\mathbf{r}, \mathbf{r}')$  is the single-particle density matrix, and  $K^{\alpha\beta}$  is a representation of the single-particle density matrix in the biorthogonal duals of the NGWFs. ONETEP achieves linear-scaling by enforcing strict localization of the NGWFs and by optimizing the density kernel and NGWFs with further localization constraints. The *in situ* optimization of the NGWFs allows a small number of orbitals to be used while maintaining accuracy comparable to plane-wave DFT codes. The NGWFs are themselves written in a basis of localized periodic cardinal sine (psinc) functions.<sup>51</sup> Systematic improvement in the accuracy of the calculation is achieved through reduction of the psinc grid spacing, which is analogous to the convergence of the kinetic energy cutoff in traditional plane-wave DFT codes. The NGWFs are initialized using an in-built pseudoatomic solver, which self-consistently solves the Kohn–Sham equations for isolated atoms with spherical confinement constraints.<sup>52</sup> Recently, a reformulation of ensemble DFT,<sup>53</sup> written in terms of NGWFs, has been implemented in ONETEP,<sup>54</sup> which allows calculations on metallic systems comprising thousands of atoms.

The ONETEP program is parallelized using a distributed memory model based on the MPI library, as described

previously.<sup>28,29</sup> Recently, the code has been extended to support hybrid MPI/OpenMP parallelism, which allows it to take advantage of state-of-the-art high performance computing architectures with hundreds of thousands of cores, for example, using shared memory within a node and distributed memory between nodes.<sup>55</sup> By integrating the DDEC/c3 method into the ONETEP code, it allows both the DFT calculation and the AIM analysis to be applied to much larger systems than has previously been possible, because the size of the system is no longer limited by the amount of memory residing on a single compute node. This approach also avoids the need to compile and run separate programs and provides the opportunity to compute AIM properties, such as atomic volumes, and use them later in the DFT calculation.<sup>56,57</sup>

In a previous paper,<sup>5</sup> we have implemented within ONETEP a charge derivation scheme based largely on the DDEC/c2 method.<sup>26</sup> IH reference densities were generated internally at run time using the pseudoatomic solver and the same exchange-correlation functional and pseudopotentials as the full DFT calculation. The resulting reference densities were conditioned to the material of interest using the approach described in eq 18, and the optimal reference density weighting of 3/14 was used.<sup>3</sup> Henceforth, we refer to this implementation as DDEC/cc2 (“conditioned c2”) to distinguish it from the DDEC/c2 and DDEC/c3 methods of Manz and Sholl. The DDEC/cc2 scheme was aimed at deriving transferable charges for use in flexible biomolecular force fields. We demonstrated on a benchmark set of 25 small, organic molecules that the DDEC/cc2 scheme gives excellent agreement with the DDEC/c3 scheme as implemented in the CHARGEMOL code.<sup>5,58</sup>

Despite the success of the DDEC/cc2 scheme in describing organic molecules, it does lack several features of the DDEC/c3 scheme that we have outlined in the previous sections, and therefore, as we will show later, it fails to satisfactorily describe NACs in dense materials with short bond lengths and diffuse atoms. To overcome these limitations, we have now implemented the full DDEC/c3 method in ONETEP. Namely, (a) constraints enforcing exponential decay of the atomic densities, (b) improvements to the c2 reference density profiles enforcing eq 15, and (c) all-electron density partitioning have been added. The DDEC/c3 constraint of eq 15 requires the initial reference densities to decay monotonically with increasing distance from the atomic center, a property not obeyed by valence (pseudo)densities derived from the norm-conserving pseudopotentials that are typically used in ONETEP. Meanwhile the generation of reference electron densities for f block ions using the PBE exchange-correlation functional and the pseudoatomic solver can fail to converge in certain cases. Therefore, in what follows, we employ the c2 reference density library supplied with the CHARGEMOL package.<sup>58</sup> These reference densities have been computed with the conductor-like polarizable continuum model in Gaussian 09<sup>59</sup> using a universal Gaussian basis set with relativistic corrections and are available for all elements in the periodic table up to atomic number 109.<sup>26</sup> The reference densities are modified in ONETEP, as described in eqs 16 and 17, to obey eq 15. Since the reference densities are now all-electron, we follow the CHARGEMOL scheme of treating systems with valence-only densities, which is to augment the latter with a set of atom-centered core densities. These core reference densities are taken to be the summed density of the corresponding Kohn–Sham DFT (core) orbitals of the neutral reference ion,  $n_A^{\text{c2}}(z_A, r_A)$ . For example, a Ga atom with 18 core electrons would have a core reference density



**Figure 1.** Outline of the DDEC/c3 implementation in ONETEP. One DDEC charge cycle involves steps A  $\rightarrow$  E. Blue boxes indicate computation that is performed in parallel with results stored separately on each node. Red boxes indicate global quantities which require node communication and whose results are stored in a distributed fashion across all nodes. In step A (Deposit), the updated weighting factors from the previous iteration for each atom distributed over all nodes are deposited onto their respective promolecular density grids, which are stored in distributed memory. In step B (Extract), for each atom, a small atom-local sample of the distributed promolecular density sufficient to encompass the largest spherical shell is extracted to the corresponding node in order to compute the weighting factors.

including the summed  $1s^2 2s^2 2p^6 3s^2 3p^6$  orbital densities of the neutral Ga reference atom. In practice, these core reference densities are also supplied by the CHARGEMOL package.<sup>58</sup> Since our DDEC treatment is now effectively all-electron, the valence-core separation scheme with core correction as described in section 2.1 is employed.

A single DFT calculation is performed on the system to obtain the ground-state electronic density, which is then processed to extract the DDEC NACs. The first step of the new DDEC/c3 implementation involves core charge derivation by using the valence-core separation scheme in order to separate

the core and valence densities, with the former fitted using the standard ISA scheme (eqs 9 and 11) and subject to the 2.0 Bohr<sup>-1</sup> exponential decay constraint. Next, the full DDEC/c3 procedure is performed on the total electronic density. In the first charge cycle, the radial Hirshfeld reference densities are initialized as neutral atomic densities. The implementation of the procedure described in section 2.1 for each DDEC/c3 charge cycle is summarized in Figure 1. Parallelism is implemented over groups of atoms local to each node, following the MPI parallelization strategy of the code for atoms.<sup>28</sup> At the start of each cycle, the promolecular weighting

factors, of the form  $n(\mathbf{r}) = \sum_{k,B} n_B(r_B)$ , are updated by depositing the relevant radial densities from each node into the distributed density grids (step A). As with previous DDEC implementations,<sup>3,5,26</sup> radial densities are stored on atom centered sets of equally spaced radial shells up to a predefined maximum cutoff radius  $r_{\max}$ . Meanwhile, the distributed densities are stored on Cartesian grids of the same spacing as that used by the DFT calculation. The Cartesian density is stored as 2D simulation slabs divided among the available nodes. This method ensures that computational cost scales linearly with the number of atoms for each DDEC iteration, because the promolecular densities and weighting factors for each point need only be computed and stored once, to be extracted when required, and the number of Cartesian grid points to be processed per atom is independent of system size.

Next, each node iterates over the groups of atoms {A} local to them. For each atom A,  $w^{\text{c3}}(\mathbf{r})$ ,  $Y^{\text{avg}}(\mathbf{r})$ , and  $w^{\text{IH}}(\mathbf{r})$  are extracted from the distributed storage into Cartesian grid boxes of size commensurate with the largest DDEC radial shell  $r_{\max}$  which are local to individual nodes (step B). The valence iterator procedure (step C) updates the radial densities and weighting factors  $w_A^{\text{ISA}}(r_A)$ ,  $w_A^{\text{IH}}(r_A)$ ,  $Y_A^{\text{avg}}(r_A)$ , and  $w_A^{\text{c3}}(r_A)$  as follows. First, the partitioned atomic density  $n_A^{i+1}(r_A)$  is updated using eq 9, and the atomic valence population  $N_A^{\text{val},i+1}$  is also calculated. The ISA weighting factor,  $w_A^{\text{ISA},i+1}(r_A)$  is then derived as the spherical average of the partitioned atomic density as in eq 11. In practice, this is calculated directly from eq 9 by reducing each vector  $\mathbf{r}$  on the right-hand side into  $r_A$  on the left-hand side. This direct spherical averaging is employed for all radial densities. Next, the conditioned reference density  $Y_A^{\text{avg},i+1}(r_A)$  is calculated using eq 18. The quantities  $\langle Y^{\pm(1/2)}(\mathbf{r}) \rangle_A$ , which are required in eq 23, are also computed during this step. The initial DDEC weighting factor  $\sigma_A^{j+1}(r_A)$  is then computed using eq 19, followed by density reshaping (minimization of eq 28) in order to transform  $\sigma_A^{j+1}(r_A)$  into the final DDEC weighting factor  $w_A^{\text{c3},i+1}(r_A)$ . Last, the IH reference weighting  $w_A^{\text{IH},i+1}(r_A)$  is updated using eq 10 with the new atomic population  $Q_A^{i+1}$  as computed earlier. Except for the distributed memory parallelization, this iterative procedure is analogous to that described by Manz and Sholl.<sup>3</sup>

Computational efficiency is achieved because the size of the Cartesian grid for each atom is much smaller than the global molecular density and independent of system size; thus, the valence iterator needs to loop over a number of Cartesian grid points that is proportional to the number of atoms. The distributed nature of the molecular density storage allows us to access large-scale systems without being hindered by memory constraints inherent in the original CHARGEMOL package, where density arrays are stored on a single node with OpenMP shared memory parallelism implemented via multithreaded loops over Cartesian grid points.<sup>58</sup>

Density reshaping of  $\sigma_A(r_A) \rightarrow w_A^{\text{c3}}(r_A)$  (eq 25) is performed at the end of each valence iterator procedure using the same computational procedure as CHARGEMOL,<sup>3</sup> by iteratively minimizing  $G_A(r_A)$  (eq 28) with an initial guess of  $G_A^0(r_A) = \sigma_A(r_A)$ . For each reshaping iteration  $j$ , the exponential decay constraint (eq 21) is enforced numerically for each consecutive shell as

$$G_A^j(r_A) = \min[G_A^j(r_A), G_A^j(r_A - \Delta r_A) e^{-\eta_A(r_A)\Delta r_A}] \quad (31)$$

starting from  $r_A = \Delta r_A > 0$ , where  $\Delta r_A$  is the (uniform) DDEC shell spacing,  $G_A(r_A)$  is then updated as

$$\Delta_A^{j+1} = \frac{\int_0^\infty 4\pi r_A^2 [\sigma_A(r_A) - G_A^j(r_A)] dr_A}{\int_0^\infty 4\pi r_A^2 \sqrt{\sigma_A(r_A)} dr_A} \quad (32)$$

$$G_A^{j+1}(r_A) = G_A^j(r_A) + \Delta_A^{j+1} \sqrt{\sigma_A(r_A)} \quad (33)$$

and this process is repeated until  $\Delta_A$  falls below a particular threshold for every atom A. The value of  $\lambda_A$  is then computed via eq 74 of reference 3. Two DDEC charge cycles are required for a complete update of the Cartesian and spherically averaged radial densities. Therefore, in practice, no density reshaping is applied for the first three DDEC cycles, where  $\lambda_A$  is set to 1 and  $G_A(r_A) = \sigma_A(r_A)$ .

### 3. COMPUTATIONAL PARAMETERS

ONETEP calculations were performed using the PBE exchange-correlation functional.<sup>60</sup> Interactions between electrons and nuclei were described by norm-conserving pseudopotentials.<sup>61</sup> NGWFs were initialized as orbitals obtained from solving the Kohn–Sham equation for isolated atoms.<sup>52</sup> One NGWF was used for H; nine for V, Fe, Mo, Pd and Pt (initialized as five d, one s, and three p orbitals); and four for all other elements used in the current study. The NGWFs were expanded as a psinc basis set<sup>51</sup> with an equivalent plane-wave cutoff energy of at least 800 eV, which means that the electron density is stored on a Cartesian grid of spacing 0.25 Bohr or smaller. The localization radii of the NGWFs were 9.0 Bohr or higher. Except for bulk supercells, where periodicity was required, calculations were performed with the cutoff Coulomb approach to avoid electrostatic interactions between molecules and their periodic images.<sup>62</sup> The ground states of the metallic systems Pt<sub>55</sub>CO, Mo<sub>2</sub>C, and Pd<sub>3</sub>V were found via self-consistent minimization of the Helmholtz free energy.<sup>53,54</sup> For these calculations, an electronic smearing of  $k_B T = 0.1$  eV was employed.  $k$ -point sampling was performed at the  $\Gamma$ -point only, though the cell sizes were large enough that the product of the number of  $k$ -points and the unit cell volume always exceeded around 4000 Å<sup>3</sup>. The only spin-polarized calculation was the DFT calculation of the myoglobin (Mb) complex with oxygen, which was treated as an open-shell singlet as described previously.<sup>63</sup>

DDEC/c3 analysis was performed as described in the previous section. For all calculations, a uniformly spaced radial grid of 100 shells with  $r_{\max} = 5$  Å corresponding to a constant shell spacing of  $\Delta r_A = 0.05$  Å was used to store the spherically averaged radial densities, while  $\gamma$  is set to the optimal value of 3/14.<sup>3</sup> The DDEC NACs were considered converged when the valence population and partitioned atomic densities, for each atom and  $r_A$ , converged below  $1 \times 10^{-5} e$  and  $1 \times 10^{-5} e/\text{Bohr}^3$ , respectively, for four consecutive iterations. For comparison, DDEC/cc2 calculations have been performed for selected systems, and in these cases, the methods are as described previously.<sup>5</sup> The root-mean-square error (RMSE) in the Coulombic potential of the DDEC NACs, compared with that of the underlying QM calculation, is given by

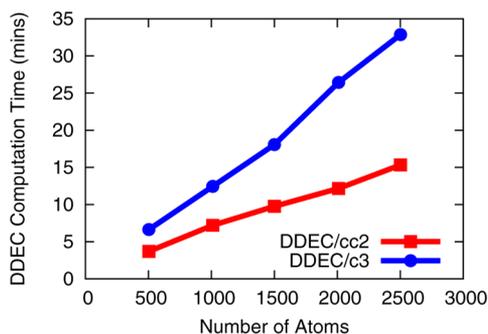
$$\text{RMSE} = \left( \frac{1}{N_{\text{Grid}}} \sum_{\mathbf{r} \in \text{Grid}} [V^{\text{DFT}}(\mathbf{r}) - V^{\text{DDEC}}(\mathbf{r}) - \langle V^{\text{DFT}}(\mathbf{r}) - V^{\text{DDEC}}(\mathbf{r}) \rangle_{\text{Grid}}]^2 \right)^{1/2} \quad (34)$$

where the sum is performed over all points  $i$  lying within 1.4 and 2.0 times the van der Waals radii of the nuclei,<sup>6</sup> on the same grid used to calculate the DFT electron density. The potentials were displaced by the averaged difference over the included grid points to remove the arbitrary vacuum level of the DFT potential. Another measure of the accuracy of the electrostatic properties of the DDEC NACs, is the relative RMSE (RRMSE), which is the ratio of the RMSE of the charge model to the RMSE of a null charge model in which all atomic charges are set to zero.

In order to test the robustness of linear-scaling DFT calculations with ONETEP, and the dependence of the results on  $k$ -point sampling, a small number of calculations were repeated using the plane-wave DFT code CASTEP. The CASTEP simulations used a 1000 eV plane-wave cutoff energy, the same pseudopotentials as the ONETEP calculations, and an electronic smearing of  $k_B T = 0.1$  eV. A  $3 \times 3 \times 3$   $k$ -point grid was used, though the convergence with respect to the number of  $k$ -points was validated as described in the text. DDEC charge analysis of the electron density was performed using the CHARGEMOL code in these cases.<sup>58</sup>

#### 4. VALIDATION

The implementation of the DDEC/c3 methodology in ONETEP is aimed at large scale chemical analysis, and so, we begin by verifying that the implementation is linear-scaling with the number of atoms in the system. Figure 2 shows the

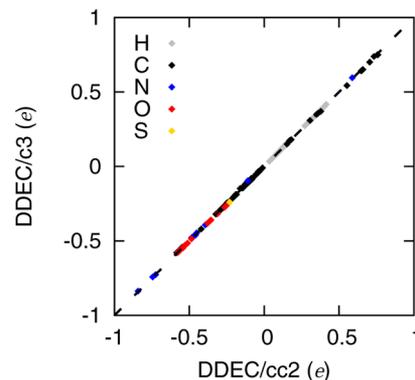


**Figure 2.** Computational time for DDEC postprocessing calculations of bulk water on 160 Intel Sandy Bridge cores. The electron density is stored on Cartesian grids of spacing 0.20 Bohr. The average DDEC/c3 NAC on the O atoms is  $-0.804e$  (compared to  $-0.783e$  for DDEC/cc2).

computational postprocessing time for DDEC/cc2 and DDEC/c3 charge analyses, using periodic boxes of bulk water of increasing size, up to around 2500 atoms, taken from our previous study.<sup>5</sup> As expected, the computational cost of both methods scales linearly with system size, thus allowing routine analysis of systems comprising thousands of atoms. The DDEC/c3 method is more expensive than DDEC/cc2 but even for the largest system studied here has a computation time of only around 30 min on 160 cores.

The average DDEC/c3 NAC on the O atom of the water molecules in the bulk DFT calculations discussed above is  $-0.80e$ , compared to the DDEC/cc2 NAC of  $-0.78e$ . The difference of  $0.02e$  between the two charge analysis schemes is larger than expected, as the DDEC/cc2 has been shown to perform well for small organic molecules.<sup>5</sup> To investigate this difference further, we compare the newly implemented DDEC/c3 charge analysis scheme in ONETEP with the DDEC/cc2

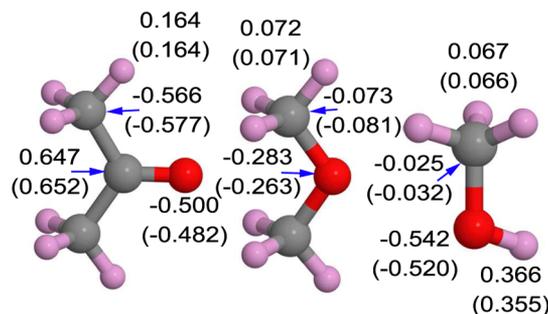
scheme for a set of 25 small, neutral organic molecules in vacuum. As shown in Figure 3, the correlation between the



**Figure 3.** Correlation between DDEC/c3 and DDEC/cc2 charges in ONETEP for all atoms in the 25 molecule benchmark set.<sup>5</sup> The RMSD between the charge sets is  $0.003e$ .

DDEC/c3 and DDEC/cc2 charges is very good and the root-mean-square deviation (RMSD) between the charge sets is just  $0.003e$ , which indicates that the new functionality in DDEC/c3 does not affect small molecules in vacuum for which DDEC/cc2 already performs well. Thus, there appears to be a larger difference between the DDEC/c3 and DDEC/cc2 NACs in bulk liquid simulations than in vacuum.

To investigate this difference between the DDEC/c3 and DDEC/cc2 NACs in more detail, we have performed separate DFT calculations of 267-molecule bulk liquid supercells of three molecules from the test set (methanol, acetone and dimethyl ether). The liquids were equilibrated under periodic boundary conditions at room temperature and pressure, using standard equilibration protocols, and the OPLS classical force field<sup>64</sup> in the boss program.<sup>65</sup> The last snapshots from the Monte Carlo simulations were used as input to the DFT calculation in ONETEP. Figure 4 shows the DDEC/c3 and



**Figure 4.** DDEC/c3 NACs of three liquids, acetone, dimethyl ether, and methanol, averaged over 251 molecules. DDEC/cc2 NACs are shown in parentheses for comparison.

DDEC/cc2 NACs averaged over all molecules in the simulation cell. Although the NACs are qualitatively similar, there are differences of up to around  $0.02e$  on some atoms, which is in excess of those seen in Figure 3. Table 1 shows the average dipole moments of the molecules in the liquid phase computed from the DDEC NACs. The liquids are more strongly polarized in the DDEC/c3 scheme, and we attribute the difference to the enforcement of exponentially decaying electron tails, which is expected to have more of an effect in bulk systems than in

**Table 1. Dipole Moments (*D*) of Molecules in Bulk Liquid (l) and Gas Phase (g) Simulations from NACs<sup>a</sup>**

		ONETEP/DDEC/ cc2	ONETEP/DDEC/ c3	experiment <sup>66</sup>
methanol	g	1.75 (1.64)	1.77 (1.64)	1.70
	l	1.85 (1.79)	1.93 (1.87)	
acetone	g	3.21 (3.10)	3.23 (3.10)	2.88
	l	3.12 (3.08)	3.32 (3.26)	
dimethyl ether	g	1.16 (1.24)	1.19 (1.24)	1.30
	l	1.41 (1.54)	1.49 (1.60)	

<sup>a</sup>Values in parentheses include atomic dipoles.

vacuum, and should act to improve the electron density partitioning in these systems. A useful advantage of the atoms-in-molecule formalism is that the partitioned electron density may be expanded to include higher order atomic multipoles. Table 1 also shows the average dipole moments of the molecules including atomic dipole (the first moment of the atomic electron density) contributions. The correction to the dipole moments obtained by including atomic dipoles is relatively small for these systems (close to 0.1*D*), indicating that a point charge model of these condensed phase systems is a reasonable one. The liquid-phase DDEC/c3 molecular dipole moments are ~10–25% larger than the measured gas phase dipole moments,<sup>66</sup> which is similar to factors that are typically used to scale gas phase charges to account for polarization in condensed phase media.<sup>67,68</sup>

Next, we investigate the performance of the ONETEP code in the description of two crystalline systems for which we expect there to be a greater difference between DDEC/c3 and DDEC/cc2 charges. Table 2 shows the NACs of bulk Mo<sub>2</sub>C in the

**Table 2. NACs of Bulk Mo<sub>2</sub>C in the PBCN Phase<sup>a</sup>**

	ONETEP ONETEP/DDEC/cc2	ONETEP ONETEP/DDEC/c3	CASTEP CHARGEMOL	VASP <sup>3</sup> CHARGEMOL
Mo	N/A	0.524	0.522	0.569
C	N/A	-1.047	-1.045	-1.139

<sup>a</sup>The first row of the table gives the software used for the DFT calculation and the second, the software used for charge analysis.

PBCN phase, with a structure identical to that used in a previous study.<sup>3</sup> The electronic structure was computed using a range of DFT software and the charge analysis performed using both ONETEP and the CHARGEMOL codes. The first point to note is that the previously described DDEC/cc2 charge analysis fails to converge for Mo<sub>2</sub>C. In this case, the calculation terminates because a charge of  $< -6e$  is assigned to the C atoms. This is expected, since short Mo–C bond lengths give large electron density overlap between atoms, and similar convergence failures have been observed using the DDEC/c2 method and the Mo<sub>2</sub>C(110) surface with an adatom.<sup>3</sup> In contrast, the DDEC/c3 method converges to a reasonable result for this system. Also shown are charges computed using CHARGEMOL, with the CASTEP plane-wave DFT code used for the underlying electronic structure calculation. Increasing the *k*-point sampling in CASTEP from a  $3 \times 3 \times 3$  *k*-point mesh to  $9 \times 9 \times 9$  changes the charges by just 0.001*e*. The ONETEP/DDEC/c3 charges are nearly indistinguishable from those using the CASTEP code. This indicates not only that there is excellent agreement between the ONETEP and CHARGEMOL implementations of the DDEC/c3 charge analysis method but also that the ONETEP calculation is

converged with respect to *k*-point sampling and basis set size. We have also compared the computed NACs with previous calculations performed in VASP, which used the projector augmented wave method. The good agreement between the two approaches indicates that the NACs are relatively insensitive to the method used to treat the core electrons. Finally, Table 3 compares the computed NACs for bulk Pd<sub>3</sub>V,

**Table 3. NACs of Bulk Pd<sub>3</sub>V<sup>a</sup>**

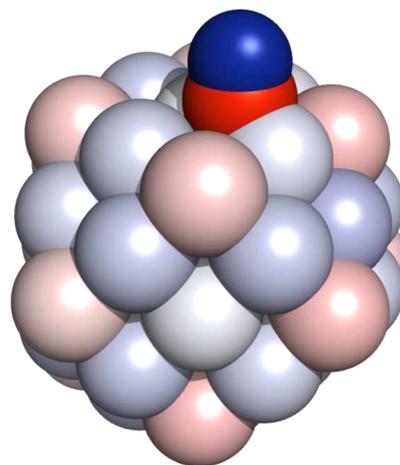
	ONETEP ONETEP/DDEC/cc2	ONETEP ONETEP/DDEC/c3	VASP <sup>3</sup> CHARGEMOL
Pd	N/A	-0.13	-0.10
V	N/A	0.39	0.31

<sup>a</sup>The first row of the table gives the software used for the DFT calculation and the second, the software used for charge analysis.

for which DDEC/c2 has been shown previously to assign unintuitive charges.<sup>3</sup> The electronegativities of Pd and V are 2.20 and 1.63, respectively, and so we expect V to be positively charged in this alloy. Indeed, the DDEC/c3 charges converge to a result that is in good agreement with previous data,<sup>3</sup> while DDEC/cc2 again fails to converge.

## 5. RESULTS

**5.1. Platinum Nanoparticle.** Dissociation of carbon monoxide (CO) on metal surfaces is a fundamental process in many catalytic processes, such as in the cleaning of automotive exhausts.<sup>69</sup> Nanoparticles can play an important role in CO oxidation, not only because their small size maximizes the catalytic surface area but also because quantum confinement effects can modify the catalytic properties of the material. Figure 5 shows a 55-atom Pt cuboctahedral nano-



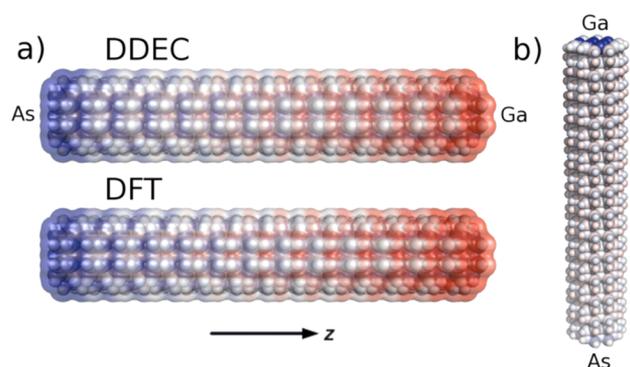
**Figure 5.** NACs for CO adsorbed on a Pt<sub>55</sub> nanocluster. Color scale ranges from +0.22*e* (red) to -0.22*e* (blue).

particle with a CO molecule adsorbed on one of the (100) facets. The geometry has been fully optimized in ONETEP, using ensemble DFT to account for partial occupancies of states near the Fermi level. The chosen particle is small enough that quantum size effects are expected to increase catalytic activity over the bulk material<sup>69,70</sup> but large enough that a range of adsorption sites are available. A typical computational investigation into the activity of Pt<sub>55</sub> would examine CO oxidation pathways at a range of adsorption sites and use the

DDEC/c3 method and other electronic structure analysis techniques to elucidate the mechanisms of catalytic rate enhancement. Such an understanding of how the nanoparticle characteristics (size, shape, surface structure, composition) affect the catalytic activity is of great industrial importance. Here, we take a simpler approach and examine the NACs of a single snapshot of the PtCO complex. In Figure 5, the atoms are colored by the DDEC/c3 NACs. The polarity of the Pt<sub>55</sub> cluster is relatively low, although there is some charge transfer from the corner to the edge sites, which may affect adsorption energies and subsequent oxidation steps of CO. The net charge transfer between the adsorbed CO molecule and the nanoparticle is very small (0.01 e). However, the catalytic surface does have the effect of substantially increasing the polarity of the C–O bond—the DDEC/c3 NACs on CO are  $\pm 0.22e$ , compared with  $\pm 0.08e$  for the isolated molecule. This observation of increased polarity on the Pt surface is in good agreement with a previous Mulliken analysis of Pt–CO complexes.<sup>71</sup>

**5.2. Polar Nanorods.** Semiconductor nanostructures are an interesting class of material, which show promise in a range of applications, such as light-emitting diodes and solar cells.<sup>72,73</sup> A contributing factor to this interest is that their electronic and optical properties differ from bulk structures of the same composition, and they can be tuned by altering the synthesis conditions.<sup>74</sup> In particular, polar semiconductor nanorods have been shown to exhibit a large dipole moment, which can affect their internal electronic structure and optical properties.<sup>75</sup> Yet the factors that determine the rods' polarity are difficult to elucidate experimentally due to the lack of precision control of the nanorod properties. Recent computational studies have extensively reviewed the dependence of the dipole moment of GaAs in the wurtzite structure on factors such as crystal symmetry, surface termination, length, and cross-sectional area.<sup>76–78</sup> By using linear-scaling DFT, as implemented in the ONETEP code, the authors were able to access size regimes comparable to experimental results. They showed that the dipole moment is strongly influenced by the surface chemistry of the nanorod, but not in an intuitive manner expected from simple charge-counting arguments. In fact, the potential difference across the nanorod is effectively limited by the size of the band gap in the structure—a concept known as *Fermi level pinning*.<sup>77,78</sup> Thus, the electrostatic properties of the rod are determined by its band gap, which itself can be altered by quantum confinement, particularly in thin rods, which necessitates large-scale QM calculations in conjunction with population analysis if we require insight into the distribution of atomic charges.

Following previous studies,<sup>76–78</sup> we have performed DFT calculations of two unrelaxed GaAs nanorods in the wurtzite structure, grown parallel to the *c*-axis. In the bulk wurtzite structure, each Ga and As atom is tetrahedrally coordinated. The polar nanorod surfaces, perpendicular to the *c*-axis, comprise one Ga-terminated surface and one As-terminated surface (Figure 6). The ends of one nanorod are left bare, and the other has dangling bonds terminated by hydrogen atoms (H-terminated). On the lateral nanorod surfaces, parallel to the *c*-axis, dangling bonds are hydrogen-terminated. The resulting rods are approximately 8.5 nm in length and 1.8 nm<sup>2</sup> in cross-sectional area and comprise approximately 1000 atoms. Following previous work in this area, we note that there is a clearly defined gap in the local density of electronic states, while the HOMO and LUMO are spatially separated at opposite ends



**Figure 6.** (a) Electrostatic potential calculated using DDEC NACs and DFT at the solvent-accessible surface of the bare polar nanorod. The color scale is clipped at  $\pm 0.5$  volts. (b) The *z*-components of the DDEC atomic dipoles. The color range is from  $-0.58$  (blue) to  $+0.58$  D (red). Ga/As terminated surfaces are also indicated.

of the rod.<sup>78</sup> Therefore, we ensured integer occupancy of the Hamiltonian eigenstates and did not apply ensemble DFT<sup>54</sup> in this case.

Table 4 shows the electrostatic properties of the two nanorods. As expected, the DFT-computed dipole moments of

**Table 4. Electrostatic Properties of Two Polar Nanorods<sup>a</sup>**

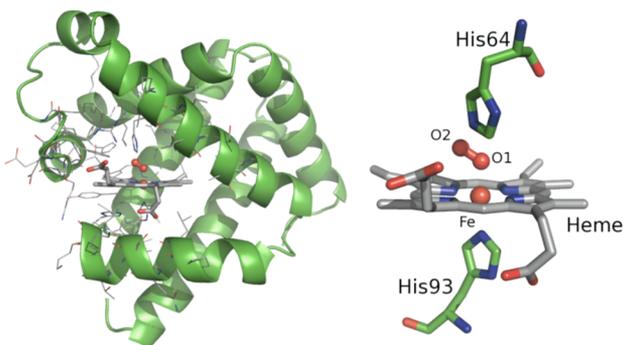
	H-terminated	bare
$\mu_{\text{DDEC}}$	-353.7 (-337.6)	177.7 (168.4)
$\mu_{\text{QM}}$	-337.6	168.4
RMSE	3.1	2.8
RRMSE	0.17	0.30

<sup>a</sup>Dipole moments (*D*) along the *z*-axis of two polar nanorods (values in parentheses include atomic dipoles), and the RMSE (kcal/mol) and RRMSE (dimensionless) of fitting the electrostatic potential using NACs.

the polar nanorods are large and the sign of the dipole is reversed upon termination of the polar surfaces with hydrogen.<sup>76–78</sup> The ability of NACs to reproduce the electrostatics of the DFT calculation is important, particularly for force field design where intermolecular interactions are often dominated by the Coulomb potential. We have computed the DDEC/c3 charges of the two nanorods from our large-scale DFT calculations. Figure 6a shows that the electrostatic potential due to the NACs is in very good agreement with that of the underlying DFT calculation, while Table 4 shows that the dipole moments of the rods are in good agreement with the DFT calculation but are overestimated in each case by approximately 5%. By computing the atomic dipoles within the atoms-in-molecule formalism (the first moment of the atomic electron density), we may visualize where they are greatest, and hence where there is a greater departure from spherical symmetry in the atomic electron densities. Figure 6b reveals that the atomic dipole moments are very small in the bulk and on the lateral surfaces but much higher on the top/bottom polar surfaces. This observation is in good agreement with previous DDEC calculations<sup>3</sup> on the SrTiO<sub>3</sub>(100) surface. The point charge model electrostatic potential is stronger than the DFT potential at the two ends of the rods. Table 4 reveals that the root-mean-square error (RMSE) in the point charge description of the electrostatic potential is around 3 kcal/mol, which is slightly higher than has been found for proteins of a

similar size ( $\sim 1.5$  kcal/mol).<sup>5</sup> The RRMSE is similar to a wide range of materials.<sup>3</sup>

**5.3. Metalloproteins.** The binding of small molecules by heme-based proteins plays a central role in the respiration process. The heme molecule is used in myoglobin (Mb) and hemoglobin proteins to, respectively, store and transport oxygen ( $O_2$ ), while its function is hindered by carbon monoxide (CO). The nature of the bonding between the central Fe atom of heme and, in particular, the  $O_2$  molecule has been the subject of much debate.<sup>79–83</sup> From a computational perspective, the calculation of the nature and energetics of binding are hindered by two major factors. First, the strong electron correlations in the localized Fe 3d orbitals strongly affect their energetic overlap with  $\pi^*$  acceptor orbitals on the  $O_2$  molecule. Second, both computation and experiment indicate that there exists substantial charge transfer from the Fe atom to  $O_2$ , an effect which is expected to stabilize the bound molecule via electrostatic interactions with the protein (in particular, with the residue His 64 shown in Figure 7).<sup>63</sup> Thus, an accurate



**Figure 7.** Heme binding site in MbO<sub>2</sub>. (Left) The full protein structure (PDB: 1A6M) is shown in green and the 1000-atom region that has been optimized in DFT as black lines. (Right) Close-up of the heme binding site showing the atomic labels used in the text.

computational description of the complex requires both a sophisticated, computationally expensive, many-body method for the Fe 3d orbitals<sup>84</sup> and a large-scale model of the protein that accurately describes long-ranged electrostatic interactions.

First, we investigate a large-scale quantum mechanical model of the myoglobin–oxygen complex using spin-polarized DFT within the generalized gradient approximation. The structure, which includes the heme molecule, ligand, and a significant portion of the protein (more than 1000 atoms), was taken from previous studies<sup>63,83</sup> and was based on the X-ray crystal structure of oxygenated sperm whale Mb (PDB: 1A6M). Table 5 summarizes the charges of the Fe atom and its neighbors, including the  $O_2$  molecule, in the DFT description of the MbO<sub>2</sub> complex, using natural population analysis (NPA),<sup>33,85–87</sup> and the DDEC/cc2 and DDEC/c3 schemes. The amount of charge transfer to the  $O_2$  molecule is reasonably consistent between the three methods, ranging from  $-0.20$  to  $-0.42e$ , and is in good agreement with a previous CASSCF/MM study of the complex.<sup>81</sup> The Fe atom is also coordinated by four nitrogen atoms in the heme molecule ( $N_{\text{heme}}$ ) and by one nitrogen atom on residue His 93. The charge transfer from Fe to N is slightly higher in the NPA method than in DDEC/c3 but within  $0.25e$  per neighbor. It has been shown previously that Fe NACs correlate well with core electron binding energies in XPS measurements of a range of solids, which supports our

**Table 5.** Charges on the Metal and  $O_2$  Molecule in a 1000-Atom Model of Myoglobin<sup>a</sup>

	GGA <sup>63</sup>			DMFT <sup>83</sup>	
	NPA	DDEC/cc2	DDEC/c3	NPA	DDEC/c3
Fe	1.13	0.45	0.58	1.13	0.48
O1	-0.17	-0.02	-0.06	-0.47	-0.42
O2	-0.25	-0.18	-0.19	-0.63	-0.56
$N_{\text{heme}}$	-0.42	-0.32	-0.34	-0.43	-0.30
$N_{\text{His93}}$	-0.38	-0.19	-0.21	-0.35	-0.20
$N_{\text{His64}}$	-0.38	-0.18	-0.16	-0.41	-0.16

<sup>a</sup>Electronic structure calculations are performed using the generalized gradient approximation (GGA) and a refined DFT+DMFT approach (DMFT).

view that the DDEC/c3 charge assignment in Fe should be reasonable in this system.<sup>3</sup>

Despite these apparently consistent data concerning the charge of the  $O_2$  molecule in MbO<sub>2</sub>, there are doubts about the approximate treatment of electron correlations in transition metal chemistry using LDA, GGA, and common hybrid exchange-correlation functionals.<sup>63,88</sup> In particular, L-edge X-ray absorption spectroscopy measurements on biomimetic heme models indicate that  $\pi$ -bonding between Fe and  $O_2$ , which is expected to mediate metal-to-ligand charge transfer, is stronger than predicted by GGA and hybrid functionals.<sup>81,82</sup> Furthermore, the stretching frequency of the O–O bond in MbO<sub>2</sub> is close to that of the unbound  $O_2^-$  ion,<sup>89,90</sup> which is inconsistent with the NACs extracted from the GGA density (Table 5). To address these contradictory results, we have applied a recently developed method that combines linear-scaling DFT with dynamical mean field theory (DFT+DMFT).<sup>91</sup> DMFT substantially improves the accuracy of strong electronic correlations in the Fe 3d subspace by taking into account quantum and thermal fluctuations, which are multireference effects not treated by conventional DFT.<sup>92</sup> DMFT also includes the Hund's exchange coupling  $J$ , which has been shown to control strong correlation effects in heme.<sup>84</sup> The ONETEP DFT+DMFT approach has been previously applied to the 1000 atom computational model of MbO<sub>2</sub> shown in Figure 7, although only Mulliken and natural population analyses were available.<sup>83</sup> Table 5 summarizes the NACs of the Fe atom, its neighbors, and the  $O_2$  molecule in the DFT+DMFT treatment of the MbO<sub>2</sub> complex at  $J = 0.7$  eV, which is a typical value used for materials containing iron.<sup>93</sup> The first point to note is that both NPA and DDEC/c3 analyses predict a much larger charge transfer to the  $O_2$  molecule in the DFT+DMFT electronic structure calculation compared with the GGA calculation. There is a small concomitant decrease in electron density on atoms surrounding the  $O_2$  binding site, and a further loss of electron density is delocalized over the 1000 atom protein and porphyrin system. The buildup of charge on the  $O_2$  molecule was previously attributed to strong  $\pi$ -bonding between the Fe 3d orbitals and the  $O_2 \pi^*$  orbitals when multireference effects are properly accounted for.<sup>83</sup> Atoms-in-molecule partitioning of the DFT+DMFT electron density confirms previous observations of strong metal-to-ligand charge transfer in MbO<sub>2</sub>,<sup>83</sup> and in fact, the net DDEC/c3 charge on the  $O_2$  molecule of  $-0.98e$  is extremely close to the chemically intuitive value of  $-1e$  that was first proposed by Weiss in the 1960s.<sup>79</sup>

## 6. CONCLUSIONS

Net atomic charges (NACs) are widely used in the chemical sciences, because they provide a convenient description of electron distribution among the atoms in a material. Multiple ways to define NACs are possible, leading to different charge assignment methods. Obviously, one should avoid methods like Mulliken charges that explicitly depend on the basis set choice and use methods that can be expressed as functionals of the electron density. AIM methods that can be expressed as functionals of the electron density include Bader, Hirshfeld, iterative Hirshfeld (IH), iterated stockholder atoms (ISA), density derived electrostatic and chemical (DDEC), and related methods. The DDEC method combines desirable properties of the IH and ISA methods. Like ISA, DDEC uses spherical averaging to minimize the atomic multipole moments to more accurately reproduce  $V(\mathbf{r})$  surrounding a material. Like IH, DDEC uses reference ion densities to improve the chemical transferability of the assigned atomic electron distributions. Although we have mostly discussed DDEC NACs in this paper, higher order multipole moments of the atomic electron densities are also available and may be used to analyze situations where the assumption of spherical atoms breaks down or to parametrize multipolar models of the true QM electrostatic potential. This expansion is similar to the ideas behind the Distributed Multipole Analysis (DMA), which has been shown to be beneficial in the description of anisotropic electronic features, including but not limited to lone pairs and  $\sigma$ -holes.<sup>35,94</sup> In addition, DDEC may in the future be useful for schemes that rely on atomic electron density distributions to compute dispersion coefficients or atomic polarizabilities.<sup>56,57</sup>

In this article, we programmed the DDEC/c3 method<sup>3</sup> into the ONETEP linear-scaling DFT code. This implementation takes advantage of ONETEP's distributed memory model to perform parallel DDEC calculations across multiple compute nodes. Both the computational cost and memory requirements scale linearly with increasing system size, allowing materials with thousands of atoms in the unit cell to be efficiently studied. This methodology adds the following improvements to the previous DDEC/cc2 implementation in ONETEP:<sup>5</sup> (a) core electrons that are modeled by an effective core potential during the DFT calculation are reinserted at the beginning of charge analysis to provide an all-electron density partitioning, (b) constraints enforce the electron density of buried atom tails to decay exponentially to improve the chemical accuracy of charge partitioning in dense materials, and (c) a precomputed library of reference ion densities is used to avoid computing reference ion densities on-the-fly and to allow access to a much wider range of chemical elements. Users of the ONETEP code will have the option of performing the original Hirshfeld electron density partitioning, or the ISA<sup>37</sup> or DDEC/c3 methods discussed in this paper. Although there have been a number of modifications to the original DDEC methodology, these have been carefully tested over a wide range of materials to ensure good performance, and thus the majority of users will not need to adjust any parameters.

The accuracy of this implementation was benchmarked using a test set of 25 small molecules, the  $\text{Mo}_2\text{C}$  solid, the  $\text{Pd}_3\text{V}$  solid, and bulk liquid supercells of water, methanol, acetone, and dimethyl ether. DDEC/c3 was much better for modeling the  $\text{Mo}_2\text{C}$  and  $\text{Pd}_3\text{V}$  solids than DDEC/cc2, which failed to converge in these cases. Moreover, results for  $\text{Mo}_2\text{C}$  and  $\text{Pd}_3\text{V}$  solids using pseudopotentials in ONETEP were similar to those

obtained using the projector augmented wave (PAW) method in VASP. DDEC/c3 and DDEC/cc2 gave essentially identical results for the 25 small molecule test set of organic molecules. Larger differences were seen for liquid methanol, dimethyl ether, and acetone, and DDEC/c3 reproduced the expected increase in polarization upon moving the molecules from the gas to the liquid phase. The latter result may be important for force field design, where the NACs should respond in an intuitive manner to electrostatic changes in their environment.<sup>5</sup>

As an example of a nanomaterial with industrial potential, we studied a pair of semiconducting GaAs nanorods comprising approximately 1000 atoms. The sign of the dipole moments of the rods depended on the surface termination, confirming previous theoretical observations that synthesis conditions are important for nanorod properties.<sup>76–78</sup> An important consideration when employing NACs in force field design is the extent to which they are able to reproduce the electrostatic properties of the underlying quantum mechanical calculation. For each of the nanorods, the DDEC/c3 NACs reproduced  $V(\mathbf{r})$  on the material's solvent accessible surface to within approximately 3 kcal/mol and reproduced the nanorod's total dipole moment to within approximately 5%. However, a large departure from spherical symmetry was observed for the electron densities of atoms on the polar surfaces of the nanorods, indicating that a multipole expansion is more suitable for these atoms.

We also studied a 55-atom Pt nanocluster with an adsorbed CO molecule. Since this material is a good electrical conductor with closely spaced energy levels surrounding the Fermi level, the electron density distribution for this material was computed using ensemble DFT. The DDEC/c3 NACs reveal that the CO molecule is strongly polarized upon adsorbing from the gas phase onto the metal surface.

As an example of a biomolecule with a heavy metal ion, for which we might expect some differences between DDEC/c3 and DDEC/cc2, we studied a 1000-atom model of oxygenated sperm whale myoglobin using previously computed generalized gradient approximation (GGA)<sup>63</sup> and dynamical mean field theory (DFT+DMFT)<sup>83</sup> electron densities. When using the GGA exchange-correlation functional, the adsorbed  $\text{O}_2$  net charge was  $-0.25$  (DDEC/c3),  $-0.20$  (DDEC/cc2), and  $-0.42$  e (NPA). In cases where DFT functionals fail to accurately describe strongly correlated electrons, multireference spin states, or self-interaction errors, we have shown that DDEC/c3 can be used to analyze the electron density from beyond-DFT methods, such as DFT+DMFT, to improve chemical understanding of such systems. Indeed, for both the natural population analysis (NPA) and DDEC/c3 methods, the adsorbed  $\text{O}_2$  net charge is approximately  $-1e$  following refinement of the localized Fe 3d subspace with dynamical mean field theory. Our DDEC/c3 results thus confirm a large metal-to-ligand charge transfer in  $\text{MbO}_2$  and help to rationalize the strong protein–oxygen electrostatic interactions that stabilize the bound complex *in vivo*.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: daniel.cole@yale.edu.

\*E-mail: tmanz@nmsu.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are grateful to Nicholas Hine (University of Cambridge), Jonah Vilseck (Yale University), and Jolyon Aarons (University of Southampton) for supplying the nanorod, bulk liquid, and PtCO structures, respectively, and to Cédric Weber (King's College London) and David O'Regan (Trinity College Dublin) for dynamical mean field theory calculations of myoglobin. T.A.M. and N.G.L. were supported by startup funds provided by New Mexico State University and were provided access to computational clusters through the Extreme Science and Engineering Discovery Environment (XSEDE) project TG-CTS100027. D.J.C. is supported by a Marie Curie International Outgoing Fellowship within the seventh European Community Framework Programme and acknowledges the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. The work of C.K.S. was supported by EPSRC grant number EP/J015059/1. Additional computational resources were provided by the Cambridge HPC Service, funded by EPSRC Grant EP/J017639/1.

## REFERENCES

- (1) Yang, Q.; Liu, D.; Zhong, C.; Li, J.-R. *Chem. Rev.* **2013**, *113*, 8261–8323.
- (2) Price, S. L.; Stone, A. J. *J. Chem. Soc. Faraday Trans.* **1992**, *88*, 1755–1763.
- (3) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2012**, *8*, 2844–2867.
- (4) Bader, R. F. W.; MacDougall, P. J.; Lau, C. D. H. *J. Am. Chem. Soc.* **1984**, *106*, 1594–1605.
- (5) Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; Jorgensen, W. L.; Payne, M. C. *J. Chem. Theory Comput.* **2013**, *9*, 2981–2991.
- (6) Watanabe, T.; Manz, T. A.; Sholl, D. S. *J. Phys. Chem. C* **2011**, *115*, 4824–4836.
- (7) Watanabe, T.; Sholl, D. S. *Langmuir* **2012**, *28*, 14114–14128.
- (8) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 4313–4323.
- (9) Jasuja, H.; Zang, J.; Sholl, D. S.; Walton, K. S. *J. Phys. Chem. C* **2012**, *116*, 23526–23532.
- (10) Vilhelmsen, L. B.; Walton, K. S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 12807–12816.
- (11) Erucar, I.; Manz, T.; Keskin, S. *Mol. Simul.* **2014**, *40*, 557–570.
- (12) Wu, X.; Huang, J.; Cai, W.; Jaroniec, M. *RSC Adv.* **2014**, *4*, 16503–16511.
- (13) Fang, H. J.; Kamakoti, P.; Zang, J.; Cundy, S.; Paur, C.; Ravikovitch, P. I.; Sholl, D. S. *J. Phys. Chem. C* **2012**, *116*, 10692–10701.
- (14) Fang, H. J.; Kamakoti, P.; Ravikovitch, P. I.; Aronson, M.; Paur, C.; Sholl, D. S. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12882–12894.
- (15) Shang, J.; Li, G.; Gu, Q.; Singh, R.; Xiao, P.; Liu, J.; Webley, P. *Chem. Commun.* **2014**, *50*, 4544–4546.
- (16) Mondal, A.; Balasubramanian, S. *J. Chem. Eng. Data* **2014**, *59*, 3061–3068.
- (17) Parkinson, G. S.; Manz, T. A.; Novotny, Z.; Sprunger, P. T.; Kurtz, R. L.; Schmid, M.; Sholl, D. S.; Diebold, U. *Phys. Rev. B* **2012**, *85*, 195450.
- (18) Antony, A.; Asthagiri, A.; Weaver, J. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12202–12212.
- (19) Montemore, M. M.; Medlin, J. W. *J. Chem. Phys.* **2012**, *136*, 204710.
- (20) Semidey-Flecha, L.; Teng, D.; Habenicht, B.; Sholl, D. S.; Xu, Y. *J. Chem. Phys.* **2013**, *138*, 184710.
- (21) Li, L. W.; Morrill, M. R.; Shou, H.; Barton, D. G.; Ferrari, D.; Davis, R. J.; Agrawal, P. K.; Jones, C. W.; Sholl, D. S. *J. Phys. Chem. C* **2013**, *117*, 2769–2773.
- (22) Rettew, R. E.; Cheng, S.; Sauerbrey, M.; Manz, T. A.; Sholl, D. S.; Jaye, C.; Fischer, D. A.; Alamgir, F. M. *Top. Catal.* **2013**, *56*, 1065–1073.
- (23) Shi, X. R.; Sholl, D. S. *J. Phys. Chem. C* **2012**, *116*, 10623–10631.
- (24) Canto, G.; Salazar-Ehuan, I.; Gonzalez-Sanchez, J.; Tapia, A.; Quijano, R.; Simonetti, S. *Int. J. Hydrogen Energy* **2014**, *39*, 8744–8748.
- (25) Kim, H. W.; Jung, J.; Han, M. N.; Lim, S.; Tamada, K.; Hara, M.; Kawa, M.; Kim, Y.; Kuk, Y. *J. Am. Chem. Soc.* **2011**, *133*, 9236–9238.
- (26) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.
- (27) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *J. Chem. Phys.* **2005**, *122*, 084119.
- (28) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *Phys. Stat. Sol. (B)* **2006**, *243*, 2489–2499.
- (29) Hine, N. D. M.; Haynes, P. D.; Mostofi, A. A.; Skylaris, C.-K.; Payne, M. C. *Comput. Phys. Commun.* **2009**, *180*, 1041–1053.
- (30) Cole, D. J.; Skylaris, C. K.; Rajendra, E.; Venkitaraman, A. R.; Payne, M. C. *EPL* **2010**, *91*, 37004.
- (31) Cole, D. J.; Rajendra, E.; Roberts-Thomson, M.; Hardwick, B.; McKenzie, G. J.; Payne, M. C.; Venkitaraman, A. R.; Skylaris, C. K. *PLoS Comp. Bio.* **2011**, *7*, e1002096.
- (32) Lever, G.; Cole, D. J.; Hine, N. D. M.; Haynes, P. D.; Payne, M. C. *J. Phys.: Cond. Matt.* **2013**, *25*, 152101.
- (33) Lee, L. P.; Cole, D. J.; Payne, M. C.; Skylaris, C.-K. *J. Comput. Chem.* **2013**, *34*, 429–444.
- (34) Lever, G.; Cole, D. J.; Lonsdale, R.; Ranaghan, K. E.; Wales, D. J.; Mulholland, A. J.; Skylaris, C.-K.; Payne, M. C. *J. Phys. Chem. Lett.* **2014**, *5*, 3614–3619.
- (35) Cardamone, S.; Hughes, T. J.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10367–10387.
- (36) Bultinck, P.; Alsenoy, C. V.; Ayers, P. W.; Carbó-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.
- (37) Lillestolen, T. C.; Wheatley, R. J. *J. Chem. Phys.* **2009**, *131*, 144101.
- (38) Hirshfeld, F. *Theor. Chim. Acta* **1977**, *44*, 129–138.
- (39) Nalewajski, R. F.; Parr, R. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8879–8882.
- (40) Heidar-Zadeh, F.; Ayers, P. W.; Bultinck, P. *J. Chem. Phys.* **2014**, *141*, 094103.
- (41) Davidson, E. R.; Chakravorty, S. *Theor. Chim. Acta* **1992**, *83*, 319–330.
- (42) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. *J. Chem. Theory Comput.* **2012**, *8*, 661–676.
- (43) Van Damme, S.; Bultinck, P.; Fias, S. *J. Chem. Theory Comput.* **2009**, *5*, 334–340.
- (44) Vanpoucke, D.; Bultinck, P.; Van Driessche, I. *J. Comput. Chem.* **2013**, *34*, 405–417.
- (45) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Theory Comput.* **2013**, *9*, 2221–2225.
- (46) Bultinck, P.; Cooper, D. L.; Van Neck, D. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3424–3429.
- (47) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- (48) Velde, G. T.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84–98.
- (49) Manz, T. A.; Sholl, D. S. *J. Comput. Chem.* **2011**, *7*, 4146–4164.
- (50) Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Dieguez, O.; Payne, M. C. *Phys. Rev. B* **2002**, *66*, 035119.
- (51) Mostofi, A. A.; Haynes, P. D.; Skylaris, C.-K.; Payne, M. C. *J. Chem. Phys.* **2003**, *119*, 8842–8848.
- (52) Ruiz-Serrano, A.; Hine, N. D. M.; Skylaris, C.-K. *J. Chem. Phys.* **2012**, *136*, 234101.
- (53) Marzari, N.; Vanderbilt, D.; Payne, M. C. *Phys. Rev. Lett.* **1997**, *79*, 1337–1340.
- (54) Ruiz-Serrano, A.; Skylaris, C.-K. *J. Chem. Phys.* **2013**, *139*, 054107.
- (55) Wilkinson, K.; Hine, N. D. M.; Skylaris, C.-K. *J. Chem. Theory Comput.* **2014**, *10*, 4782–4794.

- (56) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.
- (57) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, 236402.
- (58) Manz, T. A.; Gabaldon Limas, N. Chargemol program for performing DDEC analysis. Available from <http://ddec.sourceforge.net> (accessed September 2013).
- (59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.01; Gaussian Inc.: Wallingford, CT, 2009.
- (60) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (61) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.
- (62) Hine, N. D. M.; Dziedzic, J.; Hayne, P. D.; Skylaris, C.-K. *J. Chem. Phys.* **2011**, *135*, 204103.
- (63) Cole, D. J.; O'Regan, D. D.; Payne, M. C. *J. Phys. Chem. Lett.* **2012**, *3*, 1448–1452.
- (64) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (65) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700.
- (66) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 93rd ed.; CRC Press: Boca Raton, FL, 2012; pp 9.45–9.51.
- (67) Udier-Blagović, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *108*, 16264–16270.
- (68) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**, *10*, 2802–2812.
- (69) Dobrin, S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12122–12129.
- (70) Li, L.; Larsen, A. H.; Romero, N. A.; Morozov, V. A.; Glinsvad, C.; Abild-Pedersen, F.; Greeley, J.; Jacobsen, K. W.; Nørskov, J. K. *J. Phys. Chem. Lett.* **2013**, *4*, 222–226.
- (71) Smith, G. W.; Carter, E. A. *J. Phys. Chem.* **1991**, *95*, 2327–2339.
- (72) Tessler, N.; Medvedev, V.; Kazes, M.; Kan, S.; Banin, U. *Science* **2002**, *295*, 1506–1508.
- (73) Huynh, W. U.; Dittmer, J. J.; Alivisatos, A. P. *Science* **2002**, *295*, 2425–2427.
- (74) El-Sayed, M. A. *Acc. Chem. Res.* **2004**, *37*, 326–333.
- (75) Li, L.; Alivisatos, A. P. *Phys. Rev. Lett.* **2003**, *90*, 097402.
- (76) Avraam, P. W.; Hine, N. D. M.; Tangney, P.; Haynes, P. D. *Phys. Rev. B* **2011**, *83*, 241402.
- (77) Avraam, P. W.; Hine, N. D. M.; Tangney, P.; Haynes, P. D. *Phys. Rev. B* **2012**, *85*, 115404.
- (78) Hine, N. D. M.; Avraam, P. W.; Tangney, P.; Haynes, P. D. *J. Phys. Conf. Ser.* **2012**, *367*, 012002.
- (79) Weiss, J. J. *Nature* **1964**, *203*, 182–183.
- (80) Goddard, W. A., III; Olafson, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2335–2339.
- (81) Chen, H.; Ikeda-Saito, M.; Shaik, S. *J. Am. Chem. Soc.* **2008**, *130*, 14778–14790.
- (82) Wilson, S. A.; Kroll, T.; Decreau, R. A.; Lundberg, M.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.* **2013**, *135*, 1124–1136.
- (83) Weber, C.; Cole, D. J.; O'Regan, D. D.; Payne, M. C. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 5790–5795.
- (84) Weber, C.; O'Regan, D. D.; Hine, N. D. M.; Littlewood, P. B.; Kotliar, G.; Payne, M. C. *Phys. Rev. Lett.* **2013**, *110*, 106402.
- (85) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066–4073.
- (86) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (87) Glendenning, E. D.; Badenhop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5.9; NBO 5.9 Manual*; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI. Accessed November 2013.
- (88) Scherlis, D. A.; Cococcioni, M.; Sit, P.; Marzari, N. *J. Phys. Chem. B* **2007**, *111*, 7384–7391.
- (89) Momenteau, M.; Reed, C. A. *Chem. Rev.* **1994**, *94*, 659–698.
- (90) Jones, R. D.; Summerville, D. A.; Basolo, F. *Chem. Rev.* **1979**, *79*, 139–179.
- (91) Weber, C.; O'Regan, D. D.; Hine, N. D. M.; Payne, M. C.; Kotliar, G.; Littlewood, P. B. *Phys. Rev. Lett.* **2012**, *108*, 256402.
- (92) Georges, A.; Kotliar, G.; Krauth, W.; Rozenberg, M. *J. Rev. Mod. Phys.* **1996**, *68*, 13–125.
- (93) Tomczak, J. M.; Haule, K.; Kotliar, G. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3243–3246.
- (94) Stone, A. J. *J. Chem. Phys. Lett.* **1981**, *83*, 233–239.