# Class IV Charge Model for the Self-Consistent Charge Density-Functional Tight-Binding Method

**Jaroslaw A. Kalinowski,[†,‡] Bogdan Lesyng,*[,†,‡] Jason D. Thompson,[§] Christopher J. Cramer,*[,§] and Donald G. Truhlar*[,§]**

*Department of Biophysics, University of Warsaw, Zwirki i Wigry 93, 02-089 Warsaw, Poland, Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw, Pawinskiego 5a, 02-106 Warsaw, Poland, and Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431*

*Received: October 30, 2003; In Final Form: January 21, 2004*

Class IV charges obtained using charge model 3 (CM3) have been shown to provide a realistic description of molecular charge distributions, even when obtained by mapping from highly approximate semiempirical wave functions. In the present study, the CM3 approach is extended to the self-consistent charge density-functional tight-binding (SCC-DFTB) method. Before mapping, the mean-signed error in 219 electric dipole moments obtained by Mulliken analysis is −0.46 D, and the root-mean-square error is 0.72 D. After CM3 mapping, these errors are decreased to −0.001 and 0.31 D, respectively. The resulting charge model, denoted CM3/SCC-DFTB, should be very useful (i) for obtaining reliable charges for large molecules, nanostructures, and macromolecular systems and (ii) for representing solute charge distributions when computing the electrostatic potential or the electrostatic contribution to solvation free energies.

## 1. Introduction

The development of reliable methods that allow for the determination of electrostatic fields and interactions at the microscopic level (i.e., the atomic scale) as well as the mesoscopic level (i.e., the scale of macromolecules and nanostructured materials) is of key importance for practical applications in (bio)molecular, polymer, and materials studies. The linking of a microscopic, quantum mechanical description with a classical mechanical or quantum classical mesoscopic one is essential for reliable multiscale modeling strategies of complex systems, materials, and processes. For such applications, it is important to have a model for the computation of atomic charges that can reproduce or predict molecular charge distributions, especially if this can be accomplished with approaches that are sufficiently fast to be used to study large (bio)molecular and materials systems. One such quantum mechanical method is the self-consistent charge density-functional tight-binding method (SCC-DFTB).[1−4]

Conventional tight-binding (TB) formulations are based on decomposing the Born−Oppenheimer potential energy surface into the band energy and repulsive contributions.[5,6] Band structure is calculated from a Hamiltonian that is typically built from two-center contributions, and the repulsive energy is assumed to depend only on distances between atoms. It was shown by Foulkes and Haydock[7] that the TB formalism can be obtained as a first-order expansion of a density functional around some reference density, which allows for the parametrization of TB schemes using density-functional theory (DFT) functionals. This approach was used to develop the original DFTB method,[8] which is based on the selection of the reference electron density as the sum of the densities of isolated atoms. The second-order expansion of the Kohn−Sham DFT energy functional with respect to electron density fluctuations can be expressed in the form of Coulomb interactions of atomic charges, modified at short distances to account for the exchange-correlation energy, and it requires self-consistent charges (SCC). This second-order expansion has been incorporated in the SCC-DFTB method, which has been very successful for chemical, biochemical, and solid-state physics problems.[4] It can also be used as the quantum mechanical component of combined quantum mechanical/molecular mechanical (QM/MM) methods for calculating potential energy functions of large systems. In particular, the method has been developed with the PBE density functional[9] and incorporated into the CHARMM[10] and GROMOS[11] molecular mechanics packages and applied to proton-transfer processes.[12,13] The SCC-DFTB method is becoming increasingly popular for use in simulations of chemical, biochemical, and materials systems, as exemplified by several recent applications to a diverse set of problems involving semiconductors,[14−16] high-energy molecules,[17] enzyme catalysis,[18−21] protein dynamics,[22] oxidation−reduction potentials,[23] opiate spectra,[24] and carcinogen interactions with nucleobases.[25] Therefore, it is of interest to improve its ability to predict electrostatic properties.

In the present article, we develop a class IV charge model[26−28] as a way to obtain accurate partial atomic charges for large systems by a mapping starting with SCC-DFTB wave functions. Section 2 reviews the charge model, and section 3 presents its parametrization. Sections 4 and 5 contain results, discussion, and conclusions.

## 2. Theory

**2.1. Mulliken and Löwdin Population Analyses.** In density-functional theories, the electron density of an *N*-electron system is computed from a single-determinant wave function involving *N* Kohn−Sham orbitals, where in general a given orbital may

* Corresponding authors. E-mail: truhlar@umn.edu, b.lesyng@icm.edu.pl, cramer@chem.umn.edu.
† Department of Biophysics, University of Warsaw.
‡ Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw.
§ University of Minnesota.

10.1021/jp037288+ CCC: $27.50 
Published on Web 03/10/2004

appear once or twice in the set of orbitals; in the present study, we consider only closed-shell species where all orbitals appear twice. The orbitals are expanded in a set of atomic basis functions

$$\psi_i = \sum_\mu C_{\mu i}\phi_\mu \tag{1}$$

where $C_{\mu i}$ are the expansion coefficients and the summation is carried over all of the basis functions. There are many ways in which the electronic density can be split into parts corresponding to each atom. In the simplest case, each basis function is assigned to a nucleus.

Because the sum of the densities of all occupied orbitals equals the number of electrons, we can write

$$N = \sum_i^{\text{occ}} \int |\psi^2| = \sum_i^{\text{occ}} \sum_{\mu\nu} C^*_{\mu i} C_{\nu i} \int \phi^*_u \phi_\nu \tag{2}$$

which, if we introduce the overlap matrix

$$S_{\mu\nu} = \int \phi^*_\mu \phi_\nu \tag{3}$$

and the density matrix

$$P_{\mu\nu} = \sum_i^{\text{occ}} C^*_{\mu i} C_{\nu i} \tag{4}$$

can be rewritten as

$$N = \sum_{\mu\nu} P_{\mu\nu} S_{\nu\mu} = \sum_\mu (\mathbf{PS})_{\mu\mu} = \text{tr}(\mathbf{PS}) \tag{5}$$

where tr denotes a trace. The Mulliken population on atom $k$ is defined by partitioning this charge among the atoms as[29]

$$N_k = \sum_{\mu\in k} (\mathbf{PS})_{\mu\mu} \tag{6}$$

This definition of charges is not unique. We can also write

$$N = \text{tr}(\mathbf{PS}) = \text{tr}(\mathbf{SP}) = \text{tr}(\mathbf{S}^{1/2}\mathbf{P}\mathbf{S}^{1/2}) \tag{7}$$

and partitioning of the latter form into atomic contributions defines Löwdin populations:[30]

$$N_k = \sum_{\mu\in k} (\mathbf{S}^{1/2}\mathbf{P}\mathbf{S}^{1/2})_{\mu\mu} \tag{8}$$

Löwdin's symmetric orthogonalization procedure leads to orthogonal multicenter basis functions that have the largest possible sum of squared overlaps with the original nonorthogonal atomic basis functions. No element of the new basis set is privileged with respect to the others, as would be the case from Gram−Schmidt orthogonalization. Although typically the Löwdin population analysis gives better values of molecular electric dipole moments than the Mulliken one,[31,32] when these dipole moments are computed from the atomic point charges neither method is accurate enough for demanding applications, and the results typically depend quite strongly on the basis set.[32] Nevertheless, population-analysis charges can provide a starting point for more quantitative descriptions of molecular charge distributions. In particular, the Mulliken or Löwdin atomic charges can be modified to reproduce accurate molecular electric dipole moments with a correction procedure containing one or two coefficients per heteronuclear bond type. The coefficients depend on the method used to generate the electronic wave functions, the type of population analysis, and the basis set used. Such a procedure has already been proposed for a number of methods and basis sets, and the resulting charges are called Charge Model 1, 2, or 3 (CM1,[26] CM2,[31] or CM3[27,28]) charges or, in general, class IV charges. The correction procedure can be described as a mapping of the original charges to the new ones. In this study, we present a class IV model that we have developed for extracting accurate partial charges from SCC-DFTB wave functions. The resulting charges will be called CM3/SCC-DFTB charges.

**2.2. CM3 Mapping.** The essential idea of the CM3 procedure is to correct the systematic errors of individual bond dipoles by parametrizing a mapping to reproduce experimental or high-level theoretical electric dipole moments of a training set of primarily monofunctional molecules.[32] The following expression defines the mapping,

$$q_k = q_k^0 + \sum_{k'\neq k} T_{kk'} \tag{9}$$

where $q_k$ is the CM3 charge on atom $k$, $q_k^0$ is the original Mulliken or Löwdin charge ($Z_k - N_k$, where $Z_k$ is the nuclear charge), and $T_{kk'}$ is a function of the bond orders that determines the amount of charge to be transferred from atom $k'$ to atom $k$ in order to correct the systematic errors in the population analysis. The matrix elements $T_{kk'}$ are functions of the Mayer bond order[33] $B_{kk'}$; in particular, it is assumed that $T_{kk'}$ is a quadratic function of the bond order,[27,28,31]

$$T_{kk'} = D_{Z_kZ_{k'}} B_{kk'} + C_{Z_kZ_{k'}} B_{kk'}^2 \tag{10}$$

The $C$ and $D$ coefficients are determined by the optimization procedure. The Mayer bond order is given by[33]

$$B_{kk'} = \sum_{\lambda\in k} \sum_{\omega\in k'} (\mathbf{PS})_{\omega\lambda} (\mathbf{PS})_{\lambda\omega} \tag{11}$$

For single bonds, the Mayer bond orders are close to 1. For double bonds in unconjugated systems, they are close to 2. For conjugated bonds, they are close to 1.5, and, importantly, for pairs of atoms that are not bonded to one another, the Mayer bond order is close to zero. The latter ensures that charge redistribution due to CM3 mapping is a local phenomenon. Conservation of the total charge imposes the following constraints on elements of the $C$ and $D$ matrices:

$$C_{ZZ'} = -C_{Z'Z} \tag{12}$$

$$D_{ZZ'} = -D_{Z'Z} \tag{13}$$

In the original CM3 parametrization, it was not necessary to introduce atom "types". For example, one can use the same parameters for hydroxyl oxygen atoms and oxygen atoms in carbonyl groups. The same strategy is applied in this study. Thus, in our parametrization, $C_{ZZ'}$ and $D_{Z'Z}$ depend exclusively on the atomic numbers of the elements involved in the charge transfer.

All atomic partial charges are given in atomic units. All electric dipole moments are in D.

## 3. Parametrization Procedure

**3.1. Training Set.** Two sets of geometries have been used throughout this work: the ones used for the original CM3 charge model[32] and a set of geometries reoptimized in this study using the SCC-DFTB method. In the former set, the geometries were optimized using hybrid DFT methods and Hartree−Fock theory

**TABLE 1: Mean-Signed Errors (MSEs, D) and Root-Mean-Square Errors (RMSEs, D), as Compared to Experiment, of Electric Dipole Moments Calculated from Mulliken and Löwdin Charges of Various Compound Classes Using HDFT+HF and SCC-DFTB Geometries**

| | | HDFT+HF geometry | | | | SCC-DFTB | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mulliken | | Löwdin | | Mulliken | | Löwdin | |
| solute class | no. data | MSE | RMSE | MSE | RMSE | MSE | RMSE | MSE | RMSE |
| alcohols, phenol | 13 | −0.14 | 0.21 | 0.08 | 0.16 | −0.25 | 0.33 | −0.04 | 0.16 |
| aldehydes | 5 | −0.29 | 0.33 | −0.02 | 0.08 | −0.43 | 0.49 | −0.16 | 0.21 |
| aliphatic amines, aniline | 13 | −0.34 | 0.45 | −0.23 | 0.33 | −0.45 | 0.62 | −0.37 | 0.51 |
| amides and phenylurea | 16 | −0.35 | 0.37 | −0.13 | 0.16 | −0.29 | 0.35 | −0.07 | 0.20 |
| aromatic nitrogen heterocycles | 11 | −0.91 | 1.15 | −0.78 | 0.98 | −0.94 | 1.20 | −0.82 | 1.03 |
| bifunctional H, C, N, and O compounds | 12 | −0.34 | 0.58 | −0.21 | 0.44 | −0.40 | 0.75 | −0.29 | 0.62 |
| carboxylic acids | 9 | −0.01 | 0.12 | 0.12 | 0.17 | −0.07 | 0.17 | 0.07 | 0.12 |
| compounds containing S and P | 5 | −0.52 | 1.08 | −0.30 | 0.80 | −0.46 | 1.05 | −0.33 | 0.86 |
| esters | 6 | 0.13 | 0.17 | 0.14 | 0.17 | −0.15 | 0.19 | −0.13 | 0.18 |
| ethers | 11 | −0.12 | 0.23 | 0.07 | 0.22 | −0.05 | 0.19 | 0.14 | 0.26 |
| imines | 6 | −0.60 | 0.73 | −0.42 | 0.56 | −0.71 | 0.83 | −0.54 | 0.66 |
| inorganic organic | 7 | −0.39 | 0.68 | −0.34 | 0.62 | −0.44 | 0.71 | −0.38 | 0.65 |
| ketones | 11 | −0.49 | 0.56 | −0.22 | 0.31 | −0.51 | 0.59 | −0.24 | 0.33 |
| multifunctional P compounds | 7 | −0.91 | 1.45 | −0.47 | 1.00 | −0.61 | 1.41 | −0.10 | 1.05 |
| nitriles | 12 | −1.03 | 1.08 | −0.83 | 0.88 | −1.07 | 1.12 | −0.87 | 0.92 |
| nitrohydrocarbons | 5 | 0.14 | 0.25 | 0.25 | 0.35 | 0.45 | 0.56 | 0.58 | 0.69 |
| other C, H, and N compounds | 14 | −0.98 | 1.11 | −0.82 | 0.93 | −0.97 | 1.12 | −0.81 | 0.94 |
| other C, H, and O compounds | 12 | −0.44 | 0.56 | −0.21 | 0.34 | −0.52 | 0.67 | −0.28 | 0.46 |
| other sulfur-containing compounds | 21 | −0.50 | 0.85 | −0.51 | 0.66 | −0.47 | 0.82 | −0.45 | 0.60 |
| phosphorus | 7 | −0.71 | 0.80 | −0.91 | 1.02 | −0.56 | 0.82 | −0.69 | 1.15 |
| sulfides, disulfides | 8 | −0.27 | 0.40 | −0.54 | 0.71 | −0.29 | 0.41 | −0.55 | 0.71 |
| thiols | 8 | −0.15 | 0.19 | −0.30 | 0.34 | −0.15 | 0.20 | −0.30 | 0.34 |
| all compounds | 219 | −0.45 | 0.68 | −0.33 | 0.56 | −0.46 | 0.72 | −0.33 | 0.60 |

(the latter for amides only). For simplicity, we will refer to this set simply as the HDFT+HF geometries.

The training set used in this work is a subset of the training set[32] used to parametrize the original CM3 method. The CM3 training set contains 398 data points, including 382 data points for 382 nonamide compounds and 16 data points for 8 amide compounds. (For each amide compound, a pyramidal and planar conformation are considered.) Compounds that contain Li, F, Si, Cl, and Br and all compounds that do not have singlet ground states were not considered in this study. In addition, because the SCC-DFTB method located only one minimum—the pyramidal one—for benzamide, only this conformation was taken into account in the present study. This results in a training set of 219 data points for 212 compounds (204 data points for 204 nonamide compounds and 15 data points for 8 amide compounds) for both HDFT+HF and SCC-DFTB geometries.

**3.2. Selection of Parameters.** The mapping contains $C_{ZZ'}$ and $D_{ZZ'}$ parameters that account for corrections that are either quadratic or linear in bond orders, as indicated in eq 10. In the original parametrization of the CM3 charges,[27,28] preference was given to $D_{ZZ'}$ parameters because $C_{ZZ'}$ parameters lead to increased sensitivity with respect to changes in the bond orders. It has also been observed in both the original work and here that the inclusion of too many $C_{ZZ'}$ parameters can lead to large, unphysical values for some of the parameters. For example, in one case, $C_{PH}$ was equal to −1.12, and $D_{PH}$ was equal to 1.05. Usually these large corrections largely cancel, which is numerically unstable. Therefore, a preference for $D_{ZZ'}$ parameters was also adopted in the present work, and only two $C_{ZZ'}$ parameters (C−O and P−S) were used. It was necessary to exclude the O−P parameter with SCC-DFTB because it led to overly large offsetting corrections ($C_{OP} = -0.119$ and $D_{OP} = 0.323$). Interestingly, this was not the case when the HDFT+HF geometry training set was used, which indicates a possible deficiency in the SCC-DFTB method for the prediction of accurate geometries rather than a deficiency in the charge mapping.

**3.3. C−H Bond.** Requiring the mapped charges to reproduce electric dipole moments accurately is not sufficient to obtain a unique set of parameters because more than one charge distribution can give the same dipole moment. Thus, in the first step, the $D$ parameter for the C−H pair was optimized to obtain an average charge on hydrogen atoms in benzene and ethene equal to 0.11, a value that was justified in the CM2 study.[31] The value obtained in this way is $D = -0.0335$.

**3.4. Optimization.** For each molecule, the scalar electric dipole moment $\mu$ was computed from the atomic charges, $q_k$, as

$$\mu = \sqrt{\left(\sum_k q_k x_k\right)^2 + \left(\sum_k q_k y_k\right)^2 + \left(\sum_k q_k z_k\right)^2} \qquad (14)$$

where $x_k$, $y_k$, and $z_k$ are the Cartesian coordinates of atom $k$.

Optimization of the $C_{ZZ'}$ and $D_{ZZ'}$ parameters was carried out by a nonlinear least-squares fit using a modification of the Levenberg−Marquardt algorithm. Unlike the procedure employed in construction of the CM2 and previous CM3 charges, all parameters were optimized in a single run, which can be justified by the smaller number of species for which the present parametrization was developed.

**3.5. Software.** Several software packages have been used: MMTK[34] for computational steering, SciPy[35] for carrying out least-squares fits, VMD[36] for visualization, and R[37] for statistical analysis.

## 4. Results and Discussion

**4.1. Performance of SCC-DFTB Atomic Charges.** Table 1 presents mean-signed errors (MSE's) and root-mean-square errors (RMSE's) with respect to experimental values of the electric dipole moments derived from the Mulliken charges and from the Löwdin charges without any mapping using the HDFT+HF and SCC-DFTB geometries of the training set of molecules. The SCC-DFTB Mulliken charges significantly

**TABLE 2: CM3/SCC-DFTB Parameters**

| atomic pair | parameter value |
|---|---|
| *D* Parameters | |
| C−H | −0.03350 |
| N−C | −0.05598 |
| N−H | −0.11083 |
| O−C | 0.01444 |
| O−H | 0.00516 |
| O−N | 0.07551 |
| P−C | −0.06282 |
| P−H | −0.02401 |
| P−N | 0.04052 |
| P−O | 0.04603 |
| S−C | 0.00135 |
| S−H | 0.00962 |
| S−N | 0.14454 |
| S−O | 0.06920 |
| S−P | 0.09163 |
| *C* Parameters | |
| O−C | −0.02294 |
| S−P | −0.00057 |

underestimate the electric dipole moments. On average, the magnitude of the electric dipole moment is too small by 0.46 D. The total RMS deviation is 0.72 D, which is comparable to results obtained from the AM1 and PM3 methods (0.84 and 0.94 D, respectively) for the CM2 training set.[31] However, for some classes of compounds, the RMS deviation is significantly higher, and in the worst case (multifunctional phosphorus compounds), it is equal to 1.41 D.

The performance of the Mulliken and Löwdin charges in terms of the RMS deviation deteriorates in the case of SCC-DFTB geometries in comparison to the HDFT+HF ones, although this difference is only 0.04 D.

**4.2. Mulliken versus Löwdin Charges.** The performance of the Löwdin charges in terms of both MSE and RMSE is better than that for the Mulliken charges by about 0.1 D over both sets of geometries used in this work. However, after the input charges are mapped, there is little difference in their ability to predict electric dipole moments. The mapped Löwdin and Mulliken charge rms errors are 0.314 versus 0.309, respectively, in the case of the mappings calculated at SCC-DFTB geometries.

**4.3. Parameters.** The values of the CM3/SCC-DFTB parameters for the mapping using the Mulliken charges and SCC-DFTB geometries are presented in Table 2. Several other parametrizations were tested; however, because (i) there was very little improvement when using the Löwdin charges instead of the Mulliken charges, (ii) the use of Mulliken charges instead of Löwdin charges avoids the expense associated with computing the square root of the overlap matrix, and (iii) the mapping obtained using SCC-DFTB geometries gives reasonable results when applied to molecules in HDFT+HF geometries (see Table 3), we present here only the parameters for mapping Mulliken charges for SCC-DFTB geometries.

**4.4. Performance of CM3/SCC-DFTB Atomic Charges.** The relation between the original Mulliken charges and the charges obtained with the CM3/SCC-DFTB mapping is presented in Figure 1. It can be observed that the mapping does not lead to qualitative changes in the partial charges, suggesting that the improved performance results from physically realistic adjustments in bond dipoles.

Deviations (MSEs and RMSs) with respect to experimental values for electric dipole moments derived from the CM3/SCC-DFTB mapping of the Mulliken charges for compounds in the training set, broken down by functional group, are presented in Table 3 for SCC-DFTB and HDFT+HF geometries.

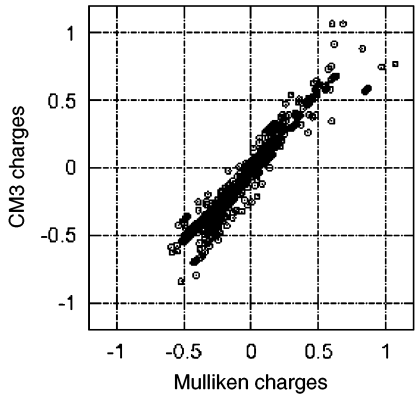


**Figure 1.** Correlation between the original Mulliken charges and CM3/SCC-DFTB charges of the molecules in the training set used in this work.

**TABLE 3: Mean-Signed Errors (MSEs, D) and Root-Mean-Square Errors (RMSEs, D), as Compared to Experiment, of Electric Dipole Moments Calculated from CM3/SCC-DFTB Charges of Various Compound Classes Using HDFT+HF and SCC-DFTB Geometries**

| | | SCC-DFTB geometry | | HDFT+HF geometry | |
|---|---|---|---|---|---|
| solute class | no. data | MSE | RMSE | MSE | RMSE |
| alcohols, phenol | 5 | 0.09 | 0.14 | 0.22 | 0.25 |
| aldehydes | 13 | 0.02 | 0.22 | 0.23 | 0.28 |
| aliphatic amines, aniline | 16 | −0.03 | 0.17 | −0.11 | 0.13 |
| amides and phenylurea | 11 | −0.29 | 0.38 | −0.25 | 0.33 |
| aromatic nitrogen heterocycles | 12 | −0.16 | 0.48 | −0.08 | 0.28 |
| bifunctional H, C, N, and O compounds | 9 | 0.21 | 0.30 | 0.25 | 0.32 |
| carboxylic acids | 5 | 0.10 | 0.34 | −0.02 | 0.24 |
| compounds containing S and P | 6 | 0.02 | 0.12 | 0.31 | 0.35 |
| esters | 11 | 0.15 | 0.27 | 0.08 | 0.22 |
| ethers | 6 | 0.11 | 0.32 | 0.25 | 0.41 |
| imines | 7 | −0.19 | 0.29 | −0.12 | 0.30 |
| inorganic organic | 11 | −0.03 | 0.20 | −0.02 | 0.21 |
| ketones | 7 | 0.09 | 0.76 | −0.36 | 0.64 |
| multifunctional P compounds | 12 | 0.12 | 0.21 | 0.17 | 0.23 |
| nitriles | 5 | 0.38 | 0.48 | 0.08 | 0.22 |
| nitrohydrocarbons | 14 | 0.10 | 0.25 | 0.08 | 0.19 |
| other C, H, and N compounds | 12 | −0.16 | 0.38 | −0.09 | 0.26 |
| other C, H, and O compounds | 21 | −0.01 | 0.40 | −0.10 | 0.54 |
| other sulfur-containing compounds | 7 | 0.21 | 0.52 | −0.09 | 0.33 |
| phosphorus | 8 | −0.12 | 0.22 | −0.13 | 0.24 |
| sulfides, disulfides thiols | 8 | −0.05 | 0.13 | −0.06 | 0.12 |
| all compounds | 219 | 0.00 | 0.31 | 0.00 | 0.30 |

The CM3/SCC-DFTB//SCC-DFTB mapping is successful in essentially eliminating the overall systematic error; it leads to MSE's of about −0.0008 D, but for some classes of compounds, the error is larger. For example, the MSE is 0.038 D in the case of nitrohydrocarbons.

The overall RMS deviation is reduced from 0.72 to 0.31 D; however, there are classes of molecules for which the model is less robust. For example, phosphorus compounds exhibit an RMS deviation of 0.76 D. Still, this is an improvement over the value of 1.41 D obtained without the mapping. It is not surprising that P has large errors because it has been found in previous work that the column of the periodic table containing N and P is notoriously hard to parametrize.

Correlations between experimental electric dipole moments and those computed using either the Mulliken charges or the CM3/SCC-DFTB//SCC-DFTB charges are presented in Figure 2. (Note that the notation //Y means "at a geometry calculated by Y".) The overall mean-unsigned errors for each method are presented in Table 4. One can see that, indeed, the CM3 mapping provides very good accuracy.

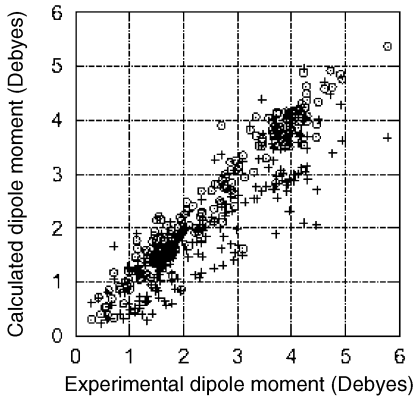


**Figure 2.** Correlation between experimental or high-level theoretical dipole moments and calculated dipole moments of the molecules in the training set used in this work. Crosses (+) correspond to dipole moments computed from Mulliken charges, and circles (○) correspond to dipole moments computed from CM3/SCC-DFTB charges.

**TABLE 4: Mean-Unsigned Errors (D), as Compared to Experiment, of Electric Dipole Moments Calculated from Various Methods and Geometries for All 219 Data Points in the Training Set**

| charge model//geometry | mean-unsigned error |
|---|---|
| Mulliken//SCC-DFTB | 0.56 |
| Mulliken//HDFT+HF | 0.52 |
| Löwdin//SCC-DFTB | 0.42 |
| Löwdin//HDFT+HF | 0.46 |
| CM3/SCC-DFTB//SCC-DFTB | 0.22 |
| CM3/SCC-DFTB//HDFT+HF | 0.22 |

## 5. Conclusions

The new class IV charge model, based on the computationally efficient self-consistent charge density-functional tight-binding method (SCC-DFTB), has been parametrized and shown to improve the accuracy of electric dipole moments computed from population analysis charges. The rms error decreases by about 0.4 D when compared with dipole moments computed from either Mulliken or Löwdin charges. The CM3 mapping parameters (Table 2) can be used for the SCC-DFTB geometries as well as for HDFT+HF ones. Accurate partial atomic charges allow one to compute very efficient representations of molecular electrostatic potentials, which have a wide variety of chemical applications.[38] The CM3/SCC-DFTB model should be very useful for large molecular systems, especially for use in combined QM/MM methods[39] or solvation models and for the development of mesoscopic potentials that are based, for example, on the generalized Born approximation.[40]

**Acknowledgment.** We thank Professor T. Frauenheim and Dr. M. Elstner for the SCC-DFTB code and consultations regarding its modifications. These studies were partially supported by the Polish State Committee for Scientific Research (4T11F 009 25) and by a Department of Defense MURI grant managed by the U.S. Army Research Office.

## References and Notes

(1) Frauenheim, T.; Weich, F.; Köhler, T.; Uhlmann, S.; Porezag, D.; Seifert, G. *Phys. Rev. B* **1995**, *52*, 11492.

(2) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185.

(3) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

(4) Frauenheim, T.; Seifert, G.; Elstner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. *Phys. Status Solidi B* **2000**, *217*, 41.

(5) Chadi, D. J. *Phys. Rev. Lett.* **1979**, *43*, 43.

(6) Goringe, C. M.; Bowler, D. R.; Hernández, E. *Rep. Prog. Phys.* **1997**, *60*, 1447.

(7) Foulkes, W. M. C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.

(8) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947.

(9) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *78*, 1396.

(10) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.

(11) Walewski, L.; Bala, P.; Lesyng, B. To be submitted for publication.

(12) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721.

(13) Meuwly, M.; Karplus, M. *J. Chem. Phys.* **2002**, *116*, 2572.

(14) Latham, C. D.; Haugk, M.; Jones, R.; Frauenheim, T.; Briddon, P. R. *Phys. Rev. B* **1999**, *60*, 15117.

(15) Rauls, E.; Staab, T. E. M.; Hajnal, Z.; Frauenheim, T. *Physica B* **2001**, *308*, 645.

(16) Rauls, E.; Frauenheim, T.; Gali, A.; Deak, P. *Phys. Rev. B* **2003**, *68*, 155208.

(17) Margetis, D.; Kaxiras, E.; Elstner, M.; Frauenheim, T.; Manaa, M. R. *J. Chem. Phys.* **2002**, *117*, 788.

(18) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721.

(19) Zhang, X. D.; Harrison, D. H. T.; Cui, Q. *J. Am. Chem. Soc.* **2002**, *124*, 14871.

(20) Smedarchina, Z.; Siebrand, W.; Fernandez-Ramos, A.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 243.

(21) Li, G. H.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 15028.

(22) Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 484.

(23) Li, G. H.; Zhang, X. D.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643.

(24) Abdali, S.; Niehaus, T. A.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Frauenheim, T.; Suhai, S.; Bohr, H. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1295.

(25) Kumar, A.; Elstner, M.; Suhai, S. *Int. J. Quantum Chem.* **2003**, *95*, 44.

(26) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(27) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 10707.

(28) Thompson J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291.

(29) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(30) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications: Mineola, NY, 1996.

(31) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(32) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117.

(33) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.

(34) Hinsen, K. *J. Comput. Chem.* **2000**, *21*, 79.

(35) www.scipy.org.

(36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(37) Ihaka, R.; Gentleman, R. *J. Comput. Graph. Stat.* **1996**, *5*, 299.

(38) *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981.

(39) See, for example, Antes, I.; Thiel, W. In *Combined Quantum Mechanical and Molecular Mechanical Models*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series 712; Americal Chemical Society: Washington, DC, 1998; p 50.

(40) Cramer, C. J.; Truhlar, D. G. *Rev. Comput. Chem.* **1995**, *6*, 1.