New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions

Jiabo Li, Tianhai Zhu, Christopher J. Cramer, and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received: August 18, 1997; In Final Form: November 7, 1997

We propose a new formalism, Charge Model 2 (CM2), to obtain accurate partial atomic charges from a population analysis of wave functions by a parametrized mapping procedure, so that the resulting charges reproduce highly accurate charge-dependent observables. The new method, which produces class IV charges, is illustrated by developing CM2 mappings of Löwdin charges obtained from semiempirical and ab initio Hartree−Fock theory and density functional theory, in particular AM1, PM3, HF/MIDI!, HF/6-31G*, HF/6-31+G*, BPW91/MIDI!, BPW91/6-31G*, B3LYP/MIDI!, and BPW91/DZVP calculations. The CM2 partial charges reproduce experimental dipole moments with root-mean-square errors that are typically a factor of 7 better than dipole moments computed from Mulliken population analysis, a factor of 3 better than dipole moments computed by Löwdin analysis, and even a factor of 2 better than dipole moments computed from the continuous electron denisty. At the HF/6-31G* and B3LYP/MIDI! levels, the new charge model yields root-mean-square errors of 0.19 and 0.18 D, respectively, for the dipole moments of a set of 211 polar molecules containing a diverse range of structures and organic functional groups and the elements H, C, N, O, F, Si, P, S, Cl, Br, and I. A comparison shows that the new charge model predicts dipole moments more accurately than MP2/cc-pVDZ calculations, which are considerably more expensive. The quality of the results is similarly good for electrostatic potentials and for the other parametrizations as well.

1. Introduction

The concept of partial atomic charge is a very powerful tool in understanding the properties of molecules.[1] It provides a simple way to characterize the electrostatic properties of electronic charge distributions for qualitative interpretations of structure and reactivity. Since electrostatic interactions are the dominant long-range contribution to intermolecular forces between molecules, the atomic charges are critical in molecular mechanics and dynamics simulations.[2] It is widely believed that the biological functions and activities of enzymes, for instance, are strongly related to the electrostatic features of their interactions.[3] Therefore, for a deeper understanding of biological processes, one needs an accurate description of the charge distribution in a biopolymer. A realistic charge model can also provide useful information for the molecular design of new drugs. Finally, the modeling of solvation effects has become a very active research area, and solvation models based on the generalized Born approximation[4,5] require high-quality partial atomic charges in order to accurately predict the electrostatic component of solvation free energy.

Unfortunately, the partial charge on an atom in a molecule cannot be unambiguously defined as it is not subject to experimental measurement. Some model must be adopted to divide the total electronic charge among the atoms in a molecule. The most widely used scheme in theoretical calculations of atomic charge is Mulliken population analysis,[6] in which the charge in a molecular orbital is partitioned into contributions associated with atomic one-electron basis functions centered at the nuclei. A similar and also widely used scheme is Löwdin population analysis, in which one partitions the charges into transformed one-electron basis functions obtained by Löwdin's symmetric orthogonalization procedure.[7] However, Mulliken charges and Löwdin charges are usually sensitive both to the

one-electron basis set and to the level of theory used for treating electron correlation. To alleviate the former problem, the so-called class III charge models have been developed to extract point charges from a quantum mechanical wave function by fitting them to a physical observable predicted from the wave function, e.g., fitting of electrostatic potentials by the ChElP[8] or ChElPG[9] schemes or fitting the calculated dipole moments.[10] Obviously, the accuracy of such a calculation depends on the quality of the wave functions. To obtain point charges that are capable of predicting physical observables accurately, high-quality wave functions are required, and in many cases such an approach is not practical. Furthermore the derivation of point charges from electrostatic potentials is often ill conditioned, especially for buried atoms. The purpose of the present work is to propose a very simple and practical scheme to obtain accurate point charges which reproduce highly accurate physical observables. Our approach follows the philosophy of class IV charge models.[11]

The classification of charge assignment models as types I, II, III, and IV is presented in ref 11, but we note here as background that traditional population analysis[6,7] yields class II charges and electrostatic fitting[8−10] yields class III charges. Class IV charges are defined by a parametrization procedure that takes as input charges from a population analysis of a wave function and maps them to reproduce charge-dependent observables obtained from experiment (or from converged quantum mechanical calculations on small molecules). Such a procedure corrects the systematic errors that can occur in the population analysis of wave functions. The method corrects simultaneously for incompleteness of the one-electron basis set, for stopping the treatment of correlation energy short of full configuration interaction, and for replacing the continuous electron density by a finite number of nuclear-centered point charges, so that

New Class IV Charge Model

J. Phys. Chem. A, Vol. 102, No. 10, 1998 **1821**

physical observables calculated from the new charges are more accurate than expectation values calculated from the original wave function. This idea has been successfully implemented in Charge Model 1 (CM1)[11] for the semiempirical model Hamiltonians AM1[12] and PM3.[13] We return to this problem here because of our conviction that even more reliable results could be obtained by mappings based on *ab initio*[14] (e.g., Hartree−Fock, HF, or Møller−Plesset second-order perturbation theory, MP2) or density functional theory[15] (DFT) calculations. In the process of exploring this possibility, we improved on the functional form of the mapping, resulting in Charge Model 2 (CM2). In the present paper we will present CM2 maps for several types of wave functions, including AM1, PM3, *ab initio* HF, and DFT.

The basic idea of CM2 mapping is simple and intuitive: although dipole moments obtained from class II charges can be poor, the error contributed from each type of bond is typically systematic, and therefore one can design systematic corrections for the bond dipole moments for each type of bond in a semiempirical way. It turns out that such a simple idea works extremely well.

A specific CM2 model requires several underlying choices: (i) the Hamiltonian or Fock operator for obtaining wave functions; (ii) the method used for population analysis of the wave functions; (iii) the functional form of the mapping; and (iv) the training-set data for parametrization of the mapping function. Section 2 is concerned with choice of (ii) and (iii), and Section 3 is concerned with choice of (i) and (iv). The CM2 mapping of Löwdin charges obtained from DFT calculations using the BPW91 functional[16,17] and MIDI! basis set[18] is discussed in detail. Section 4 provides the results and discussion. The CM2 point-charge-derived dipole moments are compared with the density-derived dipole moments obtained at a fairly high level of *ab initio* theory, namely, MP2/cc-pVDZ.[14,19]

Finally, we would like to mention a point concerning the nomenclature of the CM2 charge model: since CM2 mapping of Löwdin charges is superior to any other available scheme, when we mention CM2, it refers to CM2 mapping of Löwdin charges unless specified otherwise. At semiempirical levels that neglect differential overlap, such as AM1 and PM3, Löwdin and Mulliken charges are identical; thus there is no ambiguity there.

2. Theoretical Formalism

Löwdin Population Analysis. The two simplest choices for calculation of class II charges are Mulliken and Löwdin population analysis. We find that the errors in the dipole moments derived from Löwdin charges are typically two to four times smaller than those derived from Mulliken charges. Furthermore, in Mulliken analysis, the calculated electron population on an atomic basis function can be negative, apparently as a result of the equal partitioning of all overlap charges, which is arbitrary. Therefore, we chose Löwdin population analysis for mapping. Löwdin analysis is based on choosing a set of modified multicenter basis functions that are orthogonal while still having the greatest possible overlap with their parental atomic basis functions. Such a set of basis functions is obtained by the well-known symmetric orthogonalization introduced by Löwdin.[7] The orthogonalization procedure can be motivated by the consideration that no element of the new basis should be privileged with respect to the others (which is why the orthogonalization is referred to as "symmetric"). More rigorously, the new orthogonal basis vectors

should be as close as possible to the corresponding nonorthogonal parental ones. The Löwdin orthogonalization accomplishes this by maximizing the sum of the squares of the overlaps between the corresponding original and new orbitals under the constraint that the new orbitals are orthogonal linear combinations of the original ones.

Define **P** as the density matrix expanded in a set of nonorthogonal basis functions $\Phi = \{\phi_1, \phi_2, ..., \phi_N\}$, and let **S** be the overlap matrix with the following elements:

$$S_{ij} = <\phi_i|\phi_j> \tag{1}$$

The symmetrically orthogonalized basis functions Φ' can be obtained by the following transformation:

$$\Phi' = \Phi \mathbf{S}^{-1/2} \tag{2}$$

Each new basis function is still associated formally with the center k of the original basis function with which it has largest overlap. The Löwdin charge on an atom k, denoted as q_k, can be expressed as

$$q_k = Z_k - \sum_{i \in k}(\mathbf{S}^{1/2}\mathbf{P}\mathbf{S}^{1/2})_{ii} \tag{3}$$

where Z_k is the nuclear charge, and $\mathbf{S}^{1/2}$ is the square root of **S**. The summation runs over all basis functions associated with atom k. Mulliken population analysis bears a strong similarity to this approach, except the summation runs over $(\mathbf{PS})_{ii}$.

CM2 Mapping. We propose a very simple functional for CM2 mapping based on the following intuitive idea. It is known that bond dipole moments are approximately additive, and in many cases they remain nearly constant from molecule to molecule. Furthermore, molecular orbital calculations constitute a model chemistry with similar bond additive properties. Thus when the charges obtained from a population analysis do not give an accurate dipole moment for a molecule, we may assume that a major part of the error comes from incorrect bond dipoles. Furthermore, the errors in the bond dipoles are presumably quite systematic. The key idea of CM2 is to correct the systematic errors of bond dipoles via an empirical parametrization procedure. The above consideration leads to the following expression:

$$q_k = q_k^0 + \sum_{k' \neq k} T_{kk'}(B_{kk'}) \tag{4}$$

where q_k is the modified (CM2) charge on atom k, q_k^0 is the atomic charge obtained from Löwdin population analysis, and $T_{kk'}$ is the amount of charge transfer from atom k' to atom k to correct the bond dipole. Note that $T_{kk'}$ is a function of the bond order $B_{kk'}$. The summation in eq 4 runs over all atoms except atom k, but the form of $T_{kk'}$ given below effectively restricts contributions to atoms bonded to atom k, since other atoms have negligible bond orders $B_{kk'}$. Since total charge must remain constant, the corresponding charge transfer from k to k' should be equal in magnitude but of opposite sign; then, since $B_{kk'} = B_{k'k}$ we must have $T_{k'k} = -T_{kk'}$. We define

$$T_{kk'} = B_{kk'}(D_{kk'} + C_{kk'}B_{kk'}) \tag{5}$$

where $D_{kk'}$ and $C_{kk'}$ are two parameters which need to be determined in an empirical way for all pairwise combinations of atom types. Although any definition of bond order could be used in the above equations, here we use Mayer's bond order.[20] This is the generalization of the covalent bond index defined by Armstrong et al.[21] from the neglect of differential overlap

level to ab initio calculations. According to Mayer's analysis, the bond order between atoms k and k' is

$$B_{kk'} = \sum_{i \in k} \sum_{j \in k'} (\mathbf{PS})_{im} (\mathbf{PS})_{mj} \qquad (6)$$

There are some obvious requirements for the mapping functional: (i) If atoms k and k' have no interaction (i.e., if the bond order is zero), the charge mapped from atom k to k' or vice versa should also go to zero. (ii) $T_{kk'}$ must be a continuous function of the bond order $B_{kk'}$ so that points along reaction coordinates for chemical reactions, where chemical bonds are changing, can be handled consistently. The requirement for conservation of the total charge is naturally satisfied by enforcing the following relations:

$$C_{kk'} = -C_{k'k} \qquad (7a)$$

$$D_{kk'} = -D_{k'k} \qquad (7b)$$

In the most general case, two parameters are required for each type of atomic pair between two different kinds of atoms. (Our parameters depend only on atomic number, so in practice this means for each kind of pair of atoms with different atomic numbers. We could, of course consider different parameters for carbonyl and hydroxyl oxygens, but we found such distinctions unnecessary, which is fortunate because they could make the model ambiguous along reaction paths.) Furthermore, if only a single bond (i.e., $B_{kk'} \approx 1$) can be formed between atom k and k', then only one parameter is needed for the pair. Although one might expand eq 5 to higher powers of $B_{kk'}$, our experience so far is that it is satisfactory to have at most two parameters.

3. Parametrization

Training Sets. The primary parametrization set used here is an extension of the primary database used for the CM1 model,[11] which consisted of 186 neutral molecules. First we deleted hydroxylamine because we will include it in the secondary database. To extend the ability of the CM2 model, we added 13 phosphorus containing compounds to our training set (CM1 is not parametrized for phosphorus). To check H charges in aromatic rings and unsaturated hydrocarbons, we also added two nonpolar molecules, benzene and ethene, into the training set, for a total of 200 molecules. All molecules in the training set are listed in the Supporting Information. The experimental gas-phase dipole moments for the first 185 molecules and six of the phosphorus compounds were taken from four compilations[22] and several additional sources.[23] The precision of the experimental data is typically ±0.02 D.[22] Experimental dipole moments are lacking for seven of the phosphorus compounds, and these are obtained from high-level theoretical calculations at the B3LYP level[17,24] using the cc-pVTZ[19] basis set; in our experience, this level typically predicts dipole moments within 0.1 D. The dipole moments of the two additional hydrocarbons are zero by symmetry. We also developed a secondary training set, which is an extension of the seconary database used for the CM1 model[11] and contains 13 compounds with N−O, O−S, and P−S bonds, including hydroxylamine, formerly in the primary database.

To further examine the performance of the model, 22 ions were chosen, and ChElPG charges obtained at the MP2/cc-pVDZ level are used for comparison. Most of the partial charges for ions are available from previous work,[11] and the rest were calculated as part of the present study. The 211 polar

neutral molecules and 22 ions are quite diverse in terms of functional groups and structures.

The compounds in the primary database contain no N−O, O−S, and P−S bonds. To make up for this deficiency, a secondary database of 13 more compounds was used for the final stage of parametrization. This database contains all nine compounds of the secondary database used in the CM1 model[8] plus hydroxylamine and three additional compounds containing P−S bonds. The geometries of all the molecules in the secondary database are optimized with HF/MIDI!. Since experimental dipole moments are unavailable for all these compounds except hydroxylamine, dipole moments calculated by the high-quality theoretical B3LYP/cc-pVTZ method are used as the standard. Our experience shows that the dipole moments calculated from B3LYP/cc-pVTZ are more reliable than those calculated by MP2/cc-pVDZ. For instance, MP2/cc-pVDZ gives a dipole moment of 0.74 D for hydroxylamine, while B3LYP/cc-pVTZ//HF/MIDI! gives a value of 0.53 D, which is much closer to the experimental value of 0.59 D.

Geometries of Training Set Molecules. All molecular geometries were optimized at the HF/MIDI! level. The HF/MIDI! geometries were chosen because (1) the experimental geometries are not available for all of the molecules in the training set and (2) the MIDI! basis set[18] is specifically designed to predict reliable molecular geometries at a relatively small cost.

Basis Sets. The MIDI! basis set[18] is used for ab initio DFT and HF calculations. The DZVP basis set,[25] which is a built-in basis set in DGAUSS,[26] is also applied for DFT calculations using the BPW91 functional. For comparison, the 6-31G* basis set[14] is also used with the BPW91 functional.[16] The MIDI! basis set was originally developed by using five spherical harmonic d functions for each d shell, but some electronic structure packages only support the option of using six Cartesian d functions for each d shell. In general, "MIDI!" without further specification denotes the 5D choice, and MIDI!(6D) is the recommended name for using the six Cartesian d functions. However, in tables we will sometimes explicitly specify 5D for clarity. For 6-31G* and DZVP, we only use the 6D option, since those bases are *defined* to use Cartesian d sets. For various high-level theoretical calculations, Dunning's cc-pVDZ or cc-pVTZ basis sets are used.[19] These basis sets are defined to use 5D-type d shells.

Software. AM1 calculations were carried out by using a locally modified version of AMSOL version 6.1.[27] HF and DFT calculations were carried out by using the Gaussian-94 suite of electronic structure programs.[17]

Nonlinear Optimization. The dipole moments derived from CM2 charges are nonlinear functions of the CM2 parameters $D_{kk'}$ and $C_{kk'}$. The optimal values of the CM2 parameters are obtained by minimizing the root-mean-square (RMS) deviation between the theoretical and experimental dipole moments of the 198 polar neutral molecules in the training set using a genetic algorithm[28,29](GA). Briefly, a genetic algorithm is a nonlinear optimization technique similar to a natural selection process: survival of the fittest. The algorithm is thus a computer simulation of genetic evolution.

C−H Bond Dipole Moment. In the parametrization of the CM2 model, the molecular dipole moments are used in the error function. However, the dipole moments are not the only criterion against which to judge computed atomic charges. One obvious reason is that rather different charge distributions in a molecule can give the same dipole moment. If the dipole moment were the only criterion, the charges for certain

New Class IV Charge Model

J. Phys. Chem. A, Vol. 102, No. 10, 1998 **1823**

TABLE 1: CM2 Parameters

	AM1	PM3	HF/MIDI! 5D	HF/MIDI! 6D	HF/6-31G* [a] 6D	BPW91/MIDI! 5D	BPW91/MIDI! 6D	B3LYP/MIDI! 5D	BPW91/6-31G* 6D	BPW91/DZVP 6D
						C Parameters				
H−C	−0.020	0.003	−0.030	−0.030	−0.070	0.000	0.000	0.000	−0.080	−0.120
H−N	0.207	0.274	0.115	0.075	0.012	0.129	0.089	0.109	0.025	−0.011
H−O	0.177	0.185	0.059	−0.002	−0.052	0.083	0.000	0.086	0.000	−0.020
H−Si	−0.083	−0.021	−0.089	0.144	−0.019	0.058	−0.048	−0.069	−0.088	−0.053
H−S	0.038	0.089	−0.072	−0.042	−0.012	−0.102	−0.153	−0.098	−0.094	−0.248
C−N	0.008	−0.022	0.015	0.017	0.010	0.017	0.023	0.019	0.013	−0.001
C−O	0.026	0.025	0.081	0.084	0.053	0.071	0.069	0.079	0.060	0.030
C−Si	0.062	−0.107	0.020	0.026	0.109	0.117	0.009	0.017	0.049	0.133
C−S	−0.059	−0.033	−0.044	−0.033	−0.079	−0.023	−0.032	−0.026	−0.053	−0.015
N−O	−0.197	−0.030	0.0	0.013	0.0	0.100	0.062	0.100	0.061	0.093
						D Parameters				
H−P	0.103	0.253	−0.010	−0.036	0.069	−0.040	−0.087	−0.036	0.049	0.041
C−N	0.086	0.156	0.018	−0.027	0.049	0.032	−0.019	0.025	0.062	0.075
C−O	0.016	0.016	−0.126	−0.185	−0.063	−0.074	−0.116	−0.098	−0.028	0.008
C−F	0.019	0.025	−0.060	−0.126	−0.046	0.042	−0.010	0.017	0.049	0.058
C−P	−0.019	0.082	−0.112	−0.101	0.010	−0.028	−0.024	−0.035	0.000	0.057
C−S	0.171	0.112	0.063	0.008	0.209	0.023	0.000	0.026	0.182	0.185
C−Cl	0.027	0.117	0.020	−0.017	0.073	0.020	−0.014	0.016	0.102	0.188
C−Br	0.081	0.040	0.072	0.019	0.034	0.071	0.040	0.073	0.062	0.124
C−I	0.147	−0.032	0.177	0.148	0.163	0.138	0.120	0.146	0.166	0.161
O−P	0.088	0.181	−0.027	0.025	0.174	−0.095	−0.048	−0.095	0.080	0.205
F−P	0.252	0.244	0.009	0.095	0.320	−0.088	−0.022	−0.088	0.177	0.219
N−O	0.134	−0.043	−0.069	−0.100	−0.091	−0.136	−0.099	−0.148	−0.091	−0.100
O−S	0.0	0.056	0.050	0.075	0.185	−0.050	−0.025	−0.034	0.071	0.094
P−S	−0.080	−0.087	0.094	0.045	−0.020	0.087	0.060	0.106	−0.032	−0.007

[a] Since the 6-31G* basis set is not available for I, the MIDI! (6D) basis set is used for this element.

symmetric or nearly symmetric molecules could still be systematically in error even if a good fit were to be achieved for the dipole moments. The most clear-cut example is hydrocarbons such as CH_4, C_2H_4, C_6H_6, and so forth, since their dipole moments are zero independent of the partition of the charge between C and H. To ensure a physically meaningful charge separation in those molecules, a further constraint on the charge model is required.

A realistic charge model should ideally provide partial charges that are insensitive to basis sets and level of treatment of electron correlation. Since the additivity of bond dipole moments holds quite well in molecules, it should be possible to derive a value for each type of bond. However, except in some special cases (for instance diatomic molecules), the assignment is not unique; thus one needs to determine the bond dipole moment of at least one type of bond in some other way in order to anchor all of the others. Here we are especially concerned with C−H bond dipole moments, since they are present in almost all organic compounds. Although there are many experimental studies whose goal was to determine C−H bond dipole moments, the bond dipole, like atomic charge, is a quantity which is intrinsically not experimentally measurable, and one needs to adopt a model and to make certain assumptions. A widely used method to determine sp^2 C−H bond dipole moments is based on infrared intensity measurements, from which an average value of 0.64 D for aromatic C−H bonds was assigned,[30] which corresponds to a charge of 0.125 on H. However, this analysis may be questioned. First, high-level theoretical calculations on the triplet state of carbene at the MP2/cc-pVTZ level indicate that a charge of 0.125 on H for sp^2 C−H is an upper limit. Second, the MP2/cc-pVDZ ChElPG charge of H in benzene is 0.08, and this number also fits the quadrupole moments calculated at the same level of theory. ChElPG charges are determined in such a way that the point charges best reproduce the electrostatic potential around the molecule. It is really not possible to say which value for the charge on H in benzene is

more "realistic" with certainty, but we decided to use the ChElPG scheme to determine the C−H dipole moments, since the electrostatic potential is the critical molecular property for most interesting application areas, such as intermolecular interactions, solvation effects, and molecular dynamics.

Since the ChElPG charges on H in aromatic rings and other unsaturated compounds are actually somewhat scattered, the ChElPG charges of H were averaged over five molecules, namely, benzene, furan, pyrrole, thiophene, and ethene, from which an average value of 0.11 was obtained. The CM2 parameters for C−H are determined in such a way that the average of the H charges obtained from CM2 for these five molecules is also equal to 0.11. For mapping Löwdin charges from BPW91/MIDI! wave functions, the CM2 parameter $C_{C−H}$ turns out coincidentally to be 0.0, but in other cases, $|C_{C−H}|$ is as large as 0.08. The charge of H in benzene calculated by the CM2 mapping of the BPW91/MIDI! Löwdin charges is 0.105, which is almost exactly halfway between the ChElPG or quadrupole moment value and the value inferred from infrared intensity measurements.

Parametrization Procedure. With the C−H parameters fixed as described above, the parametrization of the remaining CM2 parameters for other atomic pairs is accomplished straightforwardly by using a genetic algorithm.[28,29] First we optimized parameters for atom pairs that are bonded in the primary database. However, simultaneous optimization of all parameters is quite inefficient. Therefore, we first optimized six parameters for the first 102 compounds, i.e., for the subset of the training set containing only H, C, N, and O. For CM2/ BPW91/MIDI!, this yielded an RMS error of 0.19 D in dipole moment; see Table 1. In the next step, those H, C, N, and O parameters were fixed, and the parameters for halogens were optimized. In the third step, all previously obtained parameters were fixed, and the parameters associated with S and Si were then optimized. In the fourth step, the parameters of bonds containing P atoms were optimized. At the final stage, the

1824 *J. Phys. Chem. A, Vol. 102, No. 10, 1998*

Li et al.

TABLE 2: Dipole Moments (D) Obtained in Various Ways from BPW91/MIDI! Wave Functions, Compared to Experiment for Representative Molecules of Each Functional Group[a]

	expt	Mulliken	Löwdin	density	CM2
alcohols					
methanol	1.70	2.59	1.55	1.55	1.65
phenol	1.45	1.99	0.96	1.33	1.24
esters, lactones					
methyl formate	1.77	1.54	1.55	1.68	1.90
γ-butyrolactone	4.27	6.17	4.16	3.80	4.56
aldehydes, ketones					
acetaldehyde	2.75	3.58	2.24	2.14	2.87
ketene	1.42	4.13	0.89	0.89	1.45
2-butanone	2.78	3.66	2.40	2.21	2.99
acids					
acetic acid	1.70	2.22	1.45	1.34	1.83
acetoacetic acid	2.30	3.52	2.07	1.98	2.18
ethers					
dimethyl ether	1.30	2.91	1.70	1.38	1.63
furan	0.66	2.01	0.61	0.34	0.45
amines					
aniline	1.53	1.27	1.39	1.74	1.66
methylamine	1.31	1.74	0.93	1.43	1.40
nitriles					
cyanoethane	4.01	5.67	2.82	3.57	4.00
benzonitrile	4.18	6.03	3.15	4.02	4.34
amides					
acetamide	3.76	3.92	3.19	3.33	3.76
formamide	3.73	3.76	3.13	3.37	3.71
imines, N-aromatics					
cis-ethyleneimine	2.06	2.79	1.63	2.16	2.36
pyrrole	1.74	0.49	1.55	2.04	1.53
multifunctional N					
3-iminofuran	1.50	1.71	0.99	0.54	1.58
aminoacetonitrile	2.64	4.17	1.96	2.22	2.70
fluorides					
methyl fluoride	1.86	3.02	1.82	1.49	2.07
fluorobenzene	1.60	2.76	1.25	0.86	1.57
chlorides					
chlorobenzene	1.69	3.34	1.48	2.15	1.68
chloroethylene	1.45	2.98	1.20	1.84	1.39
bromides					
2-bromopropane	2.21	2.49	1.39	1.95	2.01
bromoethylene	1.42	1.95	0.59	1.21	1.31
iodides					
iodobenzene	1.71	0.41	0.05	1.51	1.59
iodoethane	1.91	0.97	0.50	1.94	1.90
silicon compounds					
ethylsilane	0.81	0.21	0.15	0.91	0.75
dimethylsilane	0.71	0.35	0.07	0.86	0.78
sulfur compounds					
ethanethiol	1.58	3.07	1.52	2.00	1.38
dimethyl thioether	1.50	3.22	1.45	2.14	1.45
phosphorus compounds					
PH_3	0.57	1.95	1.29	0.89	0.88
OPH_3	1.77	0.51	1.46	1.65	1.82

[a] The HF/MIDI! geometries are used for all calculations in this article.

TABLE 3: RMS Deviation (Debye) from Experiment of BPW91/MIDI! Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups[a]

		RMS error			
type of compound	no.	Mulliken	Löwdin	density	CM2
polar H, C, N, O compounds					
alcohols, water	12	1.14	0.25	0.18	0.20
esters, lactones	8	1.13	0.41	0.48	0.19
aldehydes, ketones	16	1.48	0.62	0.69	0.22
acids	9	0.90	0.33	0.32	0.15
ethers	10	1.40	0.26	0.24	0.19
amines, ammonia	12	0.65	0.33	0.16	0.20
nitriles, HCN	17	1.69	1.15	0.39	0.15
amides	3	0.25	0.62	0.41	0.10
imines, N-aromatics	7	0.71	0.70	0.35	0.27
multifunctional N	7	1.31	1.01	0.34	0.15
subtotal	102	1.24	0.67	0.40	0.19
other polar compounds					
fluorides	31	1.25	0.44	0.69	0.25
chlorides	22	1.74	0.38	0.45	0.15
bromides	10	0.45	0.72	0.18	0.13
iodides	5	1.03	1.41	0.10	0.06
silicon compounds	6	0.53	0.59	0.15	0.07
sulfur compounds	9	2.11	0.72	0.52	0.21
phosphorus compounds	13	0.84	0.83	0.31	0.19
all polar compounds	198	1.29	0.65	0.45	0.19

[a] The HF/MIDI! geometries are used for all calculations in this article.

are compared to those obtained for several other kinds of wave functions, in particular, AM1, PM3, HF/MIDI!, HF/MIDI!(6D), HF/6-31G*, BPW91/MIDI!(6D), B3LYP/MIDI!, BPW91/6-31G*, and BPW91/DZVP. These parameters show us how the original charges obtained from Löwdin population analysis are adjusted according to eq 4. It is known that Mayer's bond order[20] corresponds well to Lewis structure bond orders, for instance, they are close to 1.0 for a single bond, 2.0 for a double bond, and 1.5 for a C−C bonds in an aromatic ring, regardless of the wave function used. The absolute values of most parameters in Table 1 are less than 0.1; thus, according to eq 5, the CM2 charges are obtained from the original Löwdin charges by only a slight adjustment. For instance, the charge for an H connected to a carbon atom remains unchanged in the BPW91/MIDI! charge model, the charge on a hydrogen bonded to a nitrogen atom is increased (made more positive) by about 0.1, and, of course, the nitrogen charge is decreased by the same amount.

General Performance of CM2. To compare results obtained from various theoretical models, we consider the BPW91/MIDI! case as an example. The 198 neutral compounds with nonzero dipole moments are classified according to functional groups, and Table 2 gives the dipole moments for a representative subset of these compounds, as calculated from Mulliken charges, from Löwdin charges, from the expectation value corresponding to the continuous three-dimensional electronic density computed from the full electronic wave function and from the CM2 charges. The results for the complete set of all molecules in the training set are given in the Supporting Information. The error in the CM2 dipoles for typical structures ranges from 0.01 to 0.30 D. The mean unsigned error for CM2/BPW91/MIDI! dipole moments for all 198 molecules is 0.15 D, and the RMS error is 0.19 D. Excluding phosphorus so that we can compare to the CM1 models published previously, the RMS error for 185 compounds remains at 0.19 D. This is a considerable improvement as compared with CM1/AM1 and CM1/PM3,[8] which each have RMS errors of 0.25 D for these 185 compounds.

secondary training set is used to parametrize N−O, O−S, and P−S bonds. At this stage, all previously obtained parameters are fixed. Four parameters were adjusted by using the secondary database: $C_{kk'}$ and $D_{kk'}$ for N−O bonds and $D_{kk'}$ for O−S and P−S bonds. To make sure that the whole procedure has reached a stable minimum in the full dimension of parameter space, the previously obtained parameters were then re-optimized. Usually this step led to no significant change in the parameters.

4. Results and Discussion

The CM2 parameters for mapping the Löwdin charges of BPW91/MIDI! wave functions are listed in Table 1, where they

New Class IV Charge Model

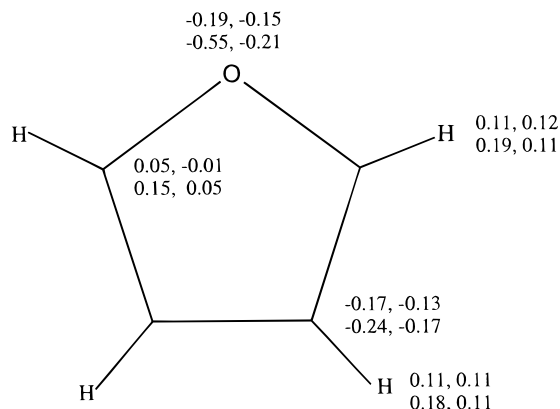*J. Phys. Chem. A, Vol. 102, No. 10, 1998* **1825**

Figure 1. Partial charges for furan. Partial charges obtained from four different methods are shown next to each atom. The first row contains CM2 and ChElPG charges, which are obtained from BPW91/MIDI! and B3LYP/cc-pVTZ wave functions, respectively. The second row contains Mulliken and Löwdin charges obtained from BPW91/MIDI! wave functions. All calculations are performed at the HF/MIDI!-optimized geometry.
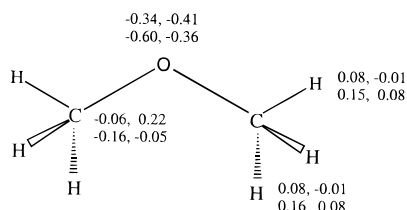

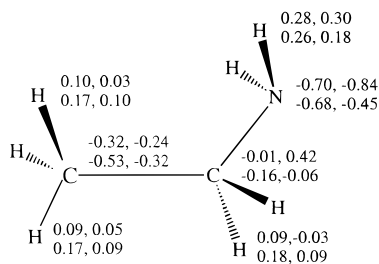**Figure 2.** Partial charges for dimethyl ether (see caption to Figure 1).

Figure 3. Partial charges for ethylamine (see caption to Figure 1).
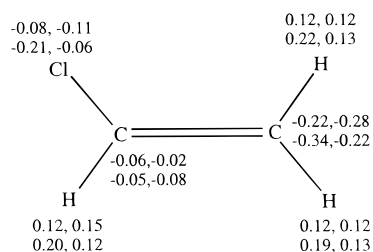

Figure 4. Partial charges for chloroethylene (see caption to Figure 1).

More detailed statistics are shown in Table 3 for the RMS errors over each type of compound. The RMS error for density-derived dipole moments is 2.4 times larger than the RMS error of CM2. Löwdin charges yield an RMS error about 3 times larger than CM2. Mulliken charges give the largest RMS error, which is about 5 times that of CM2. Partial charges obtained from CM2, Mulliken, Löwdin, and ChElPG charge models are shown for five molecules in Figures 1−5. From the five figures, one can see that CM2 charges agree better with ChElPG charges than do either Löwdin or Mulliken charges. Mulliken charges deviate most from the ChElPG values.

Comparison with MP2/cc-pVDZ Dipole Moments. One of the most striking features of the CM2 model is its accuracy
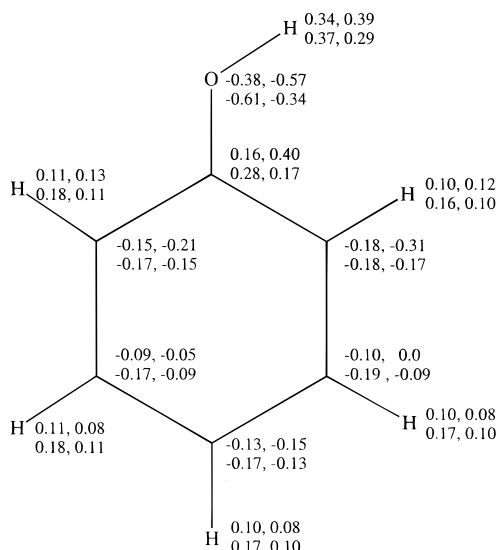

Figure 5. Partial charges for phenol (see caption to Figure 1).

TABLE 4: Comparison of CM2 Charge-Derived Dipole Moments and MP2/cc-pVDZ Density-Derived Dipole Moments with Experimental Data[a]

molecule	MP2/cc-pVDZ	CM2	exptl
methanol	1.641	1.653	1.700
phenol	1.600	1.243	1.450
water	2.038	1.847	1.850
methyl formate	1.661	1.903	1.770
acetone	2.459	3.108	2.880
formaldehyde	2.043	2.487	2.332
cyclopropanone	2.275	2.624	2.670
cis-formic acid	1.225	1.586	1.410
dimethyl ether	1.439	1.632	1.300
furan	0.491	0.448	0.661
anisole	1.312	1.612	1.380
ammonia	1.768	1.708	1.470
aniline	1.590	1.656	1.530
methylamine	1.357	1.398	1.310
hydrogen cyanide	2.752	2.762	2.985
acetonitrile	3.609	3.901	3.925
acetamide	3.527	3.759	3.760
formamide	3.606	3.712	3.730
pyrrole	2.032	1.533	1.740
cyanamide	4.145	4.035	4.320
fluoromethane	1.764	2.068	1.858
fluorobenzene	1.738	1.680	1.600
fluoroethylene	1.282	1.546	1.427
chloromethane	2.036	1.731	1.892
chlorobenzene	1.900	1.568	1.690
methanethiol	1.558	1.334	1.520
dimethyl thioether	1.708	1.453	1.500
thioformaldehyde	1.533	1.592	1.647
thiophene	0.441	0.397	0.550
hydrogen sulfide	1.388	1.029	0.970
methylsilane	0.709	0.772	0.735
PH_3	0.769	0.879	0.574
PF_3	1.340	1.169	1.025
OPF_3	1.463	1.820	1.770
CH_3PH	1.119	1.021	1.100
$(CH_3)_2PH$	1.184	0.988	1.230
mean signed error	−0.02	0.01	
mean unsigned error	0.18	0.14	
RMS error	0.21	0.17	

[a] The HF/MIDI! geometries are used for all calculations in this article. All dipole moments are in Debyes.

compared to high level density-derived dipole moment calculations. The results for a representative subset of 36 small molecules are shown in Table 4. The MP2/cc-pVDZ column

1826 *J. Phys. Chem. A, Vol. 102, No. 10, 1998*

Li et al.

TABLE 5: Selected Atomic Partial Charges for Ions

	Mulliken	Löwdin	CM2[a]	CM1/ AM1	ChElPG
N charges					
CH_2CN^-	−0.72	−0.48	−0.65	−0.68	−0.87
CH_3CNH^+	−0.42	−0.04	−0.33	−0.47	−0.34
CH_3NH^-	−0.81	−0.79	−0.96	−1.51	−1.26
$CH_3NH_3^+$	−0.67	−0.18	−0.47	−0.36	−0.34
CN^-	−0.62	−0.48	−0.68	−0.77	−0.44
$(CH_3)_2NH_2^+$	−0.66	−0.14	−0.38	−0.21	−0.14
$HCNH^+$	−0.33	0.03	−0.26	−0.43	−0.18
NH_2^-	−0.96	−1.01	−1.23	−1.75	−1.28
NO_2^-	0.11	−0.11	−0.11	−0.05	−0.16
$(CH_3)_3NH^+$	−0.67	−0.11	−0.30	−0.06	−0.01
NH_4^+	−0.69	−0.22	−0.56	−0.51	−0.77
mean signed error	−0.05	0.20	0.00	−0.11	
mean unsigned error	0.28	0.22	0.16	0.20	
RMS error	0.33	0.29	0.18	0.24	
O charges					
$CH_3COCH_2^-$	−0.70	−0.56	−0.59	−0.59	−0.78
$CH_3COHCH_3^+$	−0.45	−0.16	−0.23	−0.43	−0.35
CH_3COO^-	−0.68	−0.56	−0.61	−0.64	−0.80
CH_3O^-	−0.76	−0.70	−0.73	−0.81	−0.98
$CH_3OH_2^+$	−0.54	−0.18	−0.25	−0.40	−0.35
$CH_3OHCH_3^+$	−0.53	−0.15	−0.16	−0.28	−0.24
H_3O^+	−0.54	−0.19	−0.32	−0.50	−0.45
HO^-	−1.03	−1.03	−1.10	−1.19	−1.19
HOO^-	−0.52	−0.43	−0.50	−0.39	−0.54
$\overline{H}OO^-$	−0.71	−0.74	−0.74	−0.81	−0.75
$NO_2^{\overline{=}}$	−0.55	−0.44	−0.44	−0.47	−0.41
PO^-	−0.44	−0.25	−0.43	b	−0.39
mean signed error	−0.01	0.14	0.09	0.05[b]	
mean unsigned error	0.13	0.15	0.10	0.10[b]	
RMS error	0.15	0.17	0.13	0.12[b]	
H bonded to N					
CH_3CNH^+	0.46	0.33	0.41	0.53	0.52
CH_3NH^-	0.07	0.06	0.16	−0.02	0.22
$CH_3NH_3^+$	0.41	0.29	0.37	0.35	0.34
$(CH_3)_2NH_2^+$	0.40	0.28	0.36	0.34	0.31
$HCNH^+$	0.48	0.35	0.42	0.53	0.51
$(CH_3)_3NH^+$	0.40	0.28	0.36	0.32	0.32
NH_2^-	−0.02	0.00	0.11	0.38	0.14
NH_4^+	0.43	0.31	0.39	0.38	0.44
mean signed error	−0.02	−0.11	−0.03	0.01	
mean unsigned error	0.08	0.11	0.06	0.07	
RMS error	0.09	0.13	0.06	0.11	
H bonded to O					
$CH_3COHCH_3^+$	0.45	0.35	0.40	0.48	0.46
$CH_3OH_2^+$	0.49	0.38	0.42	0.48	0.48
$CH_3OHCH_3^+$	0.48	0.37	0.41	0.46	0.47
H_3O^+	0.51	0.40	0.44	0.50	0.48
HO^-	0.03	0.03	0.10	0.19	0.19
HOO^-	0.23	0.17	0.23	0.20	0.19
mean signed error	−0.01	−0.10	−0.05	0.00	
mean unsigned error	0.04	0.10	0.06	0.02	
RMS error	0.07	0.11	0.06	0.03	
cumulative total of atoms above					
mean signed error	−0.02	0.06	0.01	−0.01[b]	
mean unsigned error	0.15	0.15	0.10	0.11[b]	
RMS error	0.21	0.20	0.13	0.16[b]	

[a] CM2/BPW91/MIDI! [b] CM1/AM1 is not defined for P.

TABLE 6: Dipole Moment (Debye) of Ions[a]

	MP2/cc-pVDZ[b]	CM2[c]	deviation
$PO_4H_2^-$	5.346	5.226	0.12
PO^-	0.006	0.233	−0.23
CH_2CN^-	1.354	1.271	0.08
$CH_2NH_2^+$	0.040	0.116	−0.08
CH_3CNH^+	0.030	1.177	−1.15
$CH_3COCH_2^-$	2.787	3.263	−0.48
$CH_3COHCH_3^+$	1.506	1.740	−0.23
CH_3COO^-	2.721	3.417	−0.70
CH_3NH^-	2.213	2.522	−0.31
$CH_3NH_3^+$	2.199	2.173	0.03
CH_3O^-	1.389	1.807	−0.42
$CH_3OH_2^+$	1.996	1.892	0.10
$CH_3OHCH_3^+$	1.210	1.237	−0.03
CN^-	0.809	0.586	0.22
$NH_2(CH_3)_2^+$	1.420	1.480	−0.06
H_3O^+	1.585	1.781	−0.20
$HCNH^+$	1.189	0.531	0.66
HO^-	1.031	1.235	−0.20
HOO^-	2.520	2.326	0.19
NH_2^-	1.529	1.455	0.07
NO_2^-	0.596	0.633	−0.04
$NH(CH_3)_3$	0.741	0.862	−0.12
OCN^-	1.293	1.235	0.06
HS^-	0.097	0.635	−0.54
H_3S^+	1.980	1.800	0.18
$CH_3SH_2^+$	1.996	1.892	0.10
RMS error			0.36

[a] The dipole moments are calculated with respect to the nuclear charge center of the ions. [b] Density-derived dipole moments of MP2/cc-pVDZ//HF/MIDI! wave functions. [c] CM2 dipole moments calculated from BPW91/MIDI!//HF/MIDI! wave functions.

method. MP2/cc-pVDZ yields an RMS error of 0.21 D, while CM2/BPW91/MIDI! gives an RMS error of 0.17 D.

Ions. In developing charge models, ions are very challenging systems, since ions can include bonds having different bond orders and bond lengths from those found in neutral molecules. Unlike the case for neutral molecules, the dipole moment of a charged molecule depends on the choice of origin, making it quite sensitive to the geometry of molecule. Moreover, for obvious technical reasons, there are rarely experimental data available. Therefore, high-quality theoretical dipole moments are used as the standard, and the CM2 charges are also compared with ChElPG charges obtained at the MP2/cc-pVDZ//HF/MIDI! level. Table 5 shows partial charges on nitrogen, oxygen, and hydrogen atoms in ions containing H, C, N, and O and one ion containing P. In examining the trends in this table, the reader should recall that CM2/BPW9/MIDI! is relatively inexpensive compared to ChElPG/MP2/cc-pVDZ, and CM1/AM1 is extremely inexpensive.

As compared with CM1/AM1,[11] CM2/BPW91/MIDI! gives better agreement with ChElPG charges of nitrogen atoms. For charges on oxygen atoms, both methods give very similar RMS errors. For H bonded to N, CM2 yields a RMS error about half the RMS error of CM1/AM1; while for H connected to O, both CM2 and CM1/AM1 yield small RMS errors. Overall, CM2 gives a slightly better match than CM1 and a much better match than traditional population analysis with ChElPG charges obtained at the MP2/cc-pVDZ//HF/MIDI! level.

Table 6 shows the dipole moments of ions calculated from MP2/cc-pVDZ density and CM2 partial charges of BPW91/MIDI! wave functions. The dipole moments are calculated with respect to the nuclear charge centers of the ions. The statistics of 26 ions show an RMS deviation of 0.36 D between CM2 and MP2/cc-pVDZ dipole moments. For comparison with the RMS deviation of 0.36 D obtained with CM2/BPW91/MIDI!,

gives the dipole moments calculated with dipole moment operators and the continuous electronic density of MP2 wave functions. All 23 molecules listed in Table 15 of ref 11 are included. To more completely represent the diversity in the CM2 training set, additional molecules have been added, including phosphorus compounds. CM2 predicts more accurate dipole moments than the much more expensive MP2/cc-pVDZ

New Class IV Charge Model

J. Phys. Chem. A, Vol. 102, No. 10, 1998 **1827**

TABLE 7: Comparison between CM1 and CM2 Charge Models for the Primary Database

		RMS error of dipole moment (Debye)				
		CM1		CM2		
type of compound	no.	AM1 Mulliken[b]	BPW91 Löwdin	AM1 Mulliken[b]	BPW91 Mulliken	BPW91 Löwdin
alcohols, water	12	0.21	0.31	0.17	0.18	0.20
esters, lactones	8	0.24	0.35	0.18	0.28	0.19
aldehydes, ketones	16	0.18	0.20	0.17	0.65	0.22
acids	9	0.41	0.13	0.16	0.20	0.15
ethers	10	0.18	0.41	0.20	0.25	0.19
amines, ammonia	13	0.17	0.23	0.17	0.30	0.20
nitriles, HCN	17	0.17	0.16	0.21	0.40	0.15
amides	3	0.69	0.15	0.23	0.26	0.10
imines, N-aromatics	7	0.43	0.33	0.31	0.54	0.27
multifunctional N	7	0.47	0.22	0.30	0.64	0.15
subtotal (H, C, N, O)	102	0.29	0.26	0.21	0.42	0.19
fluorides	31	0.24	0.26	0.25	0.62	0.25
chlorides	22	0.18	0.19	0.15	0.54	0.15
bromides	10	0.08	0.10	0.15	0.15	0.13
iodides	5	0.10	0.12	0.15	0.11	0.06
silicon compounds	6	0.11	0.06	0.15	0.10	0.07
sulfur compounds	9	0.17	0.80	0.45	0.82	0.21
cumulative subtotal	185	0.25	0.28	0.22	0.48	0.19
phosphorus compounds	13	c	0.32	0.55	0.23	0.19
total	198		0.29	0.25	0.47	0.19

[a] The CM1 method as applied to both AM1/Mulliken charges and BPW91 Löwdin charges involves a special tanh functional for nitriles and cyanides as explained elsewhere.[8] In CM1/BPW91/MIDI!, we have also added a similar tanh functional for carbonyls. *[b]* Mulliken and Löwdin analyses are the same for AM1 because of the zero-overlap assumption. *[c]* CM1/AM1 is not defined for P.

we note that the density-derived dipole moments of BPW91/MIDI! show an RMS error of 0.50 D, while CM1/AM1 shows an error of 0.74 D, which is more than twice the error of CM2.

CM1 vs CM2 and Mulliken vs Löwdin. In evaluating the utility of the CM1 and CM2 models, we have made a comprehensive study of CM1 and CM2 mappings for charges obtained from different population analyses and different wave functions, in particular we used AM1-Mulliken, BPW91/MIDI!(Löwdin), and BPW91/MIDI!(Mulliken) charges. For this purpose we parametrized three new models specifically for this comparison; in particular we parametrized CM1/BPW91/MIDI!(Löwdin), CM2/AM1(Mulliken), and CM2/BPW91/MIDI!(Mulliken). The results for these models are compared to the original CM1/AM1 model and the standard CM2/BPW91/MIDI! model in Table 7. Note that in the text, CM1 without a parenthetical comment denotes use of the CM1 functional form with Mulliken charges, and CM2 without a parenthetical comment denotes use of the CM2 functional form with Löwdin charges, whereas if we use the CM1 functional with Löwdin charges or the CM2 functional with Mulliken charges, we append a parenthetical tag. Comparison of CM2/AM1(Mulliken) to CM1/AM1 or of CM2/BPW91/MIDI! to CM1/BPW91/MIDI!(Löwdin) shows that when both functional forms are applied to the same original initial charges, the CM2 map performs 12−32% better, even though some special C−N parameters are present in the CM1 map. (The 0.03 D difference between CM1/AM1 and CM2/AM1 may be statistically insignificant, but the 0.09 D difference between CM1 and CM2 for BPW91 Löwdin is definitely meaningful.) CM2 gives good results for nitrogen compounds in a more universal way. Then, comparing CM2/BPW91(Mulliken) to CM2/BPW91, where the latter, standard method employs Löwdin charges, we see that one obtains much better results (the error is reduced by a factor of 2.5) when mapping Löwdin charges than Mulliken charges.

Generalized Born Polarization Energy. We anticipate that one of the most important applications of the CM2 charge model

will be in estimating solute−solvent intermolecular interactions. A widely used theory for the solvation polarization energy is the generalized Born approximation, in which a solvent is represented by a continuum medium with a dielectric constant. The polarization energy G_P is given as follows[4,5]

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_k \sum_{k'} q_k q_{k'} \gamma_{kk'} \qquad (8)$$

where ϵ is the dielectric constant of the solvent, q_k is the partial charge on atom k, and $\gamma_{kk'}$ is a Coulomb integral between atoms k and k'. A good charge model used for solvation calculations should predict a physically meaningful polarization energy. However, solvation polarization energy is not measurable from experiments because free energy is not separable into electrostatic and nonelectrostatic contributions. Since ChElPG charges fit the electrostatic potential of a molecule, the polarization energy calculated from high-quality ChElPG charges should be a reasonable value. Table 8 lists the polarization energies in aqueous solution ($\epsilon = 78.3$) as obtained from various charge models by using gas phase wave functions (i.e., these are not self-consistent-reaction-field calculations). The results of CM2/BPW91/MIDI!, Löwdin, CM1/AM1, and ChElPG charges are listed. One can see that CM2 predicts polarization energies in much better agreement with those obtained from high-level-theory ChElPG charges than either Löwdin or CM1/AM1. The RMS error of CM2 is only about 42% that of CM1/AM1 and only 31% that of Löwdin charges. Such a good agreement between CM2 and high-quality ChElPG charges indicates that the CM2 model provides more realistic partial charges than CM1, and therefore we expect that it will be very useful in solvation calculations and other practical applications to intermolecular interactions. Combination of CM2 charges with the SM5 solvation model[5d,f] at the HF and DFT levels is on-going and will be reported in due course.

TABLE 8: Generalized Born Aqueous Polarization Energies Calculated from Different Charge Models for 12 Representative Molecules

molecules	polarization energy[a]			
	CM2[b]	Löwdin[c]	CM1/AM1	ChElPG[d]
ethanol	−3.47	−2.63	−3.90	−3.64
phenol	−4.36	−3.31	−5.52	−4.54
formaldehyde	−3.88	−2.04	−3.06	−3.60
methylamine	−1.87	−0.87	−2.29	−2.02
formamide	−6.32	−4.36	−5.12	−6.41
pyrrole	−2.81	−2.34	−4.36	−3.52
cyanoethane	−4.11	−1.98	−3.63	−4.08
furan	−1.55	−1.64	−3.04	−1.41
dimethyl thioether	−1.30	−1.30	−1.74	−1.50
methyl chloride	−1.12	−0.93	−1.42	−1.53
benzene	−1.60	−1.60	−2.43	−0.83
ethane	−0.36	−0.35	−0.19	−0.01
MSE[e,f]	0.03	0.81	−0.30	
MUE[e,g]	0.29	1.04	0.70	
RMS error[e]	0.37	1.21	0.88	

[a] Energy units: kcal/mol. All these polarization energies depend on the values of the Coulomb radii and $d_{kk'}$ parameters in the Coulomb integrals. In order to make a consistent comparison we used the SM5.4/U values[5d] for all calculations. Furthermore we used unrelaxed HF/MIDI! geometries, and we did not allow any of the solute electronic wave functions to relax in solution. [b] Charges are obtained by mapping Löwdin charges of BPW91/MIDI! wave functions. [c] Charges are obtained from BPW91/MIDI! wave functions by Löwdin population analysis. [d] ChElPG charges are obtained from B3LYP/cc-pVTZ//HF/MIDI! wave functions. [e] With respect to last column. [f] Mean signed error. [g] Mean unsigned error.

TABLE 9: RMS Deviation (Debye) from Experiment of BPW91/MIDI!(6D) Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups

type of compound	no.	RMS error			
		Mulliken	Löwdin	density	CM2
H, C, N, O compounds					
alcohols, water	12	1.14	0.42	0.17	0.22
esters, lactones	8	1.20	0.42	0.46	0.19
aldehydes, ketones	16	1.60	0.17	0.66	0.24
acids	9	0.96	0.18	0.30	0.14
ethers	10	1.41	0.60	0.22	0.22
amines, ammonia	13	0.62	0.19	0.17	0.20
nitriles, HCN	17	1.74	0.53	0.37	0.17
amides	3	0.26	0.33	0.38	0.15
imines, N-aromatics	7	0.69	0.39	0.34	0.28
multifunctional N	7	1.34	0.56	0.33	0.16
subtotal	102	1.28	0.40	0.38	0.21
other polar compounds					
fluorides	31	1.30	0.29	0.67	0.26
chlorides	22	1.75	0.28	0.46	0.16
bromides	10	0.31	0.38	0.18	0.13
iodides	5	1.08	1.27	0.10	0.06
silicon compounds	6	0.55	0.39	0.14	0.05
sulfur compounds	9	2.12	0.80	0.53	0.21
phosphorus compounds	13	0.75	0.50	0.33	0.23
all polar compounds	198	1.31	0.45	0.44	0.20

CM2 Mapping for Different Basis Sets and Hamiltonians. The power of the CM2 mapping is apparently universal. We examined CM2 models at semiempirical molecular orbital and ab initio HF levels, we tested different basis sets and DFT functionals, and we found in all these cases that CM2 works very well. The CM2 parameters for 10 different cases are shown in Table 1. The error statistics for seven of these models are shown in Tables 3 and 9−14, and the error statistics for the other three are in the Supporting Information. Table 9 shows the results for BPW91/MIDI!(6D). Even though we developed

TABLE 10: RMS Deviation (Debye) from Experiment of HF/MIDI! Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups

type of compound	no.	RMS error			
		Mulliken	Löwdin	density	CM2
H, C, N, O polar compounds					
alcohols, water	12	1.99	0.37	0.23	0.16
esters, lactones	8	2.19	0.50	0.20	0.18
aldehydes, ketones	16	2.60	0.13	0.24	0.16
acids	9	2.04	0.30	0.23	0.20
ethers	10	2.25	0.60	0.18	0.15
amines, ammonia	13	1.19	0.25	0.25	0.18
nitriles, HCN	17	2.91	0.80	0.16	0.16
amides	3	0.84	0.26	0.06	0.12
imines, N-aromatics	7	1.23	0.55	0.26	0.33
multifunctional N	7	2.10	0.55	0.22	0.14
subtotal	102	2.18	0.52	0.21	0.19
other polar compounds					
fluorides	31	2.53	0.61	0.15	0.21
chlorides	22	1.83	0.28	0.54	0.16
bromides	10	0.92	0.47	0.15	0.08
iodides	5	0.98	1.29	0.17	0.37
silicon compounds	6	0.41	0.61	0.12	0.09
sulfur compounds	9	1.94	0.60	0.43	0.23
phosphorus compounds	13	1.41	0.43	0.27	0.30
all polar compounds	198	2.04	0.54	0.27	0.19

TABLE 11: RMS Deviation (Debye) from Experiment of HF/MIDI!(6D) Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups

type of compound	no.	RMS error			
		Mulliken	Löwdin	density	CM2
C, N, O polar compounds					
alcohols, water	12	1.90	0.72	0.24	0.17
esters, lactones	8	2.15	0.86	0.20	0.18
aldehydes, ketones	16	2.59	0.57	0.23	0.18
acids	9	2.00	0.68	0.24	0.19
ethers	10	2.15	0.99	0.19	0.16
amines, ammonia	13	1.10	0.28	0.25	0.19
nitriles, HCN	17	2.99	0.23	0.16	0.17
amides	3	0.86	0.20	0.05	0.19
imines, N-aromatics	7	1.15	0.35	0.27	0.33
multifunctional N	7	2.14	0.41	0.18	0.23
subtotal	102	2.14	0.58	0.21	0.20
other polar compounds					
fluorides	31	2.48	0.94	0.33	0.21
sulfur compounds	9	1.97	0.72	0.46	0.27
chlorides	22	1.83	0.38	0.56	0.17
bromides	10	0.68	0.13	0.16	0.10
iodides	5	2.07	1.16	0.18	0.35
silicon compounds	6	0.46	0.42	0.12	0.18
phosphorus compounds	13	1.31	0.53	0.28	0.31
all polar compounds	198	2.04	0.64	0.30	0.21

the MIDI! basis set with the 5D option, one can see that the CM2 mappings for both the 5D and 6D options are equally good by comparing Tables 3 and 9. Table 10 shows the results of HF/MIDI!(5D) calculations. The results of HF/MIDI!(6D) are shown in Table 11. Since the MIDI! basis set is optimized at the HF level,[8] the density-derived dipole moments are considerably more accurate than those obtained from DFT calculations (see Table 3). Nevertheless, the CM2 mapping at the DFT level is slightly better than at the HF level. The DFT calculations using the B3LYP[24] functional are shown in Table 12, and the CM2 mapping gives the best results obtained from calculations using the MIDI! basis set. The CM2 mapping quality is slightly improved from the MIDI!(6D) basis set to 6-31G* (see Tables 9 and 13). Table 14 shows the results for

New Class IV Charge Model

J. Phys. Chem. A, Vol. 102, No. 10, 1998 **1829**

TABLE 12: RMS Deviation (Debye) from Experiment of B3LYP/MIDI! Dipole Moments for Groups of Compounds in the Primary Database with Various Functional Groups

type of compound	no.	RMS error			
		Mulliken	Löwdin	density	CM2
H, C, N, O polar compounds					
alcohols, water	12	1.22	0.25	0.16	0.18
esters, lactones	8	1.23	0.38	0.42	0.18
aldehydes, ketones	16	1.56	0.54	0.61	0.23
acids	9	1.02	0.27	0.27	0.14
ethers	10	1.48	0.30	0.19	0.15
amines, ammonia	13	0.67	0.32	0.15	0.17
nitriles, HCN	17	1.74	1.14	0.38	0.15
amides	3	0.27	0.58	0.36	0.17
imines, N-aromatics	7	0.74	0.68	0.32	0.28
multifunctional N	7	1.36	1.01	0.33	0.19
subtotal	102	1.30	0.65	0.36	0.18
other polar compounds					
fluorides	31	1.43	0.36	0.55	0.24
sulfur compounds	9	2.08	0.70	0.48	0.21
chlorides	22	1.78	0.35	0.49	0.15
bromides	10	0.47	0.70	0.17	0.12
iodides	5	1.10	1.43	0.10	0.06
silicon compounds	6	0.52	0.61	0.11	0.07
hydrocarbons	2	0.00	0.00	0.00	0.00
phosphorus compounds	13	0.84	0.83	0.31	0.21
all polar compounds	198	1.35	0.63	0.40	0.18

TABLE 13: RMS Deviation (Debye) from Experiment of BPW91/6-31G* Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups

type of compounds	no.	RMS error			
		Mulliken	Löwdin	density	CM2
H, C, N, O polar compounds					
alcohols	12	0.65	0.22	0.16	0.17
esters, lactones	8	0.54	0.46	0.20	0.21
aldehydes, ketones	16	0.44	0.69	0.31	0.20
acids	9	0.46	0.42	0.14	0.12
ethers	10	0.73	0.23	0.12	0.14
amines	13	0.20	0.21	0.22	0.13
nitriles	17	0.84	1.17	0.16	0.16
amides	3	0.16	0.42	0.12	0.05
imines, N-aromatics	7	0.39	0.74	0.30	0.31
multifunctional N	7	0.79	0.94	0.15	0.15
subtotal	102	0.59	0.68	0.20	0.18
other polar compounds					
fluorides	31	0.74	0.38	0.46	0.23
sulfur compounds	9	0.75	0.74	0.29	0.20
chlorides	22	0.37	0.58	0.14	0.14
bromides	10	0.45	0.30	0.08	0.13
iodides	5	0.93	1.37	0.09	0.11
silicon compounds	6	0.43	0.60	0.13	0.04
phosphorus compounds	13	1.01	0.92	0.25	0.30
all polar compounds	198	0.64	0.66	0.25	0.19

a CM2 mapping of the Löwdin charges of the HF/6-31G* wave function; this combination works very well. Mappings of similar quality are obtained from BPW91/DZVP wave functions as shown in the Supporting Information. Finally, the results of CM2 mapping of AM1 and PM3 wave functions are also shown in the Supporting Information. CM2 mappings of ab initio HF and DFT charges give better results than CM2 mappings of AM1 and PM3 charges. Tables 3 and 9−14 plus the three analogous tables in Supporting Information reveal a general trend in that the quality of the CM2 mapping depends much less strongly on the level from which the Löwdin charges are obtained than do the charges obtained by Löwdin or Mulliken analysis.

TABLE 14: RMS Deviation (Debye) from Experiment of HF/6-31G* Dipole Moments for Groups of Compounds in the Primary Neutral Database with Various Functional Groups

type of compound	no.	RMS error			
		Mulliken	Löwdin	density	CM2
H, C, N, O polar compounds					
alcohols, water	12	1.41	0.62	0.43	0.17
esters, lactones	8	1.13	0.39	0.35	0.15
aldehydes, ketones	16	1.05	0.62	0.33	0.14
acid	9	1.37	0.56	0.44	0.21
ethers	10	0.93	0.41	0.39	0.24
amines, ammonia	13	1.24	0.54	0.38	0.16
nitriles, HCN	17	0.99	0.52	0.30	0.18
amides	3	1.42	0.47	0.21	0.10
imines,N-aromatics	7	0.94	0.31	0.31	0.19
multifunctional N	7	0.98	0.45	0.33	0.13
subtotal	102	1.14	0.52	0.36	0.17
other polar compounds					
fluorides	31	1.09	0.55	0.34	0.12
chlorides	22	1.15	0.61	0.31	0.19
bromides	10	1.50	0.79	0.34	0.11
iodides	5	1.24	0.74	0.22	0.08
silicon compounds	6	1.66	0.80	0.22	0.21
sulfur compounds	9	0.79	0.60	0.29	0.09
phosphorus compounds	13	1.73	1.38	0.45	0.30
all polar compounds	198	1.22	0.66	0.35	0.17

For many purposes, such as calculations on certain negative ions and excited states, it is desirable to use diffuse functions in a basis set,[30] and we tried the CM2 mapping procedure on three basis sets containing diffuse functions, namely, 6-31+G*, 6-31++G*, and 6-31+G**. None of these basis sets yields charges as good as the basis sets without diffuse functions. This is seen in the RMS errors of dipole moments obtained from Löwdin analysis as well as in the CM2 results. It seems that basis sets with diffuse functions are not as well suited to population analysis in normal molecules as are basis sets with only tight functions, but perhaps this is the price we have to pay to get better descriptions of states with electron density in diffuse orbitals. Nevertheless, by using more *D* parameters (and less *C* parameters), we were able to get reasonable results with diffuse functions, and of the three bases mentioned above we obtained the best dipole moments from a CM2 mapping of the 6-31+G* basis. Partial charges obtained this way could be useful for calculations on excited states since, e.g., the RMS error for compounds containing H, C, N, and O is only 0.36 D, as compared to 1.08 D from the unmapped charges. However, since the results are not as good as for the basis sets without diffuse functions, the 6-31+G* parameters and error statistics are given in the Supporting Information rather than the in the printed version of the paper.

Table 15 summarizes our results for CM2/AM1, CM2/PM3, four CM2/HF models, and five CM2/DFT models. First of all we see that excellent results are obtained from HF/MIDI!, HF/6-31G*, and B3LYP/6-31G* densities as compared to other results based on unmapped densities. One can see that HF/6-31G* gives the best results. Table 16 gives CM2 dipole moments compared to those from B3LYP/cc-pVTZ//HF/MIDI!

Parametrization of N−O, O−S, and P−S Bonds. The performance obtained from BPW91/MIDI!(6D) for CM2 dipole moments is quite good. It is also interesting to note that without introducing N−O parameters, CM2 predicts a dipole moment of 0.80 D for hydroxylamine, and with inclusion of the N−O parameters obtained from the secondary training database, CM2 gives 0.65 D, which is much closer to the experimental value. Table 17 gives a summary of the results of all CM2 models for

1830 *J. Phys. Chem. A, Vol. 102, No. 10, 1998*

Li et al.

TABLE 15: Comparison of RMS Errors in Dipole Moments (D) of 198 Polar Compounds in the Primary Neutral Database

	Mulliken	Löwdin	density	CM2
HF/6-31+G*	2.56	1.34	0.44	0.41
AM1	0.84	0.84	0.49	0.25
PM3	0.94	0.94	0.43	0.23
HF/MIDI!(6D)	2.04	0.64	0.30	0.21
BPW91/MIDI!(6D)	1.31	0.45	0.44	0.20
BPW91/DZVP	0.51	0.73	0.19	0.20
BPW91/MIDI!	1.29	0.65	0.45	0.19
HF/MIDI!	2.04	0.54	0.27	0.19
BPW91/6-31G*	0.64	0.66	0.25	0.19
B3LYP/MIDI!	1.35	0.63	0.40	0.18
HF/6-31G*	1.22	0.66	0.35	0.17

TABLE 16: CM2 Dipole Moments Compared with the Density Derived Dipole Moments Obtained from B3LYP/cc-pVTZ//HF/MIDI!

molecule	dipole moments, D		
	CM2[a]	B3LYP/ cc-pVTZ	deviation
oxazole [C_3H_3NO]	2.551	2.885	−0.334
formaldoxime [CH_2NOH]	0.109	0.152	−0.043
dimethyl sulfoxide [$(CH_3)_2SO$]	4.038	3.977	0.061
dimethylsulfone [$(CH_3)_2SO_2$]	4.584	4.536	0.047
nitromethane [CH_3NO_2]	3.133	3.324	−0.191
methanesulfonamide [$CH_3SO_2NH_2$]	3.265	3.242	0.023
methanesulfonic acid [CH_3SO_3H]	2.900	2.921	−0.021
methyl methanesulfenate [CH_3SOCH_3]	0.482	0.199	0.283
methyl methanesulfinate [$CH_3S(O)OCH_3$]	2.614	2.812	−0.198
hydroxylamine [NH_2OH]	0.651	0.590	0.061
$CH_3P(O)(OCH_3)(SCH_3)$	1.448	1.488	−0.040
$CH_3P(O)(SCH_3)_2$	1.407	1.351	0.056
$OP(OCH_3)(SCH_3)_2$	1.909	1.603	0.306
RMS error			0.17

[a] CM2 dipole moments are obtained from BPW91/MIDI!(6D)//HF/MIDI! wave functions.

TABLE 17: Comparison of RMS Errors in Dipole Moments (D) of the 13 Compounds in the Secondary Database and the Total of 211 Neutral Compounds

	secondary[a]	primary + secondary[b]
AM1	0.79	0.30
HF/6-31+G* [c]	0.77	0.44
BPW91/DZVP	0.53	0.23
PM3	0.38	0.24
HF/6-31G*	0.37	0.19
BPW91/6-31G*	0.36	0.21
BPW91/MIDI!	0.19	0.19
B3LYP/MIDI!	0.19	0.18
BPW91/MIDI!(6D)	0.17	0.20
HF/MIDI!	0.13	0.19
HF/MIDI!(6D)	0.12	0.21

[a] 13 compounds. [b] 211 compounds. [c] Since 6-31+G* is not defined for iodine, we used MIDI!(6D) augmented by diffuse *s* and *p* shells with exponent 0.03 for iodine.

the secondary database. As one can see, the MIDI! basis set gives quite good results for all compounds in the secondary training set, while 6-31+G* gives the worst results for reasons discussed in the previous section.

5. Conclusion

The present paper presents a new approach to obtaining class IV charges. In this approach we introduce semiempirical parameters for pairs of atoms in the interpretation of wave functions. In this way one can obtain the best possible charges from a given level of quantum mechanics, because the semi-empirical mapping is optimized so that the physical observables calculated from the improved point charges are as accurate as possible. Furthermore, the CM2 mapping does not suffer from ill conditioning for buried atoms or large molecules, as traditional electrostatic fitting methods often do.

The new scheme we have proposed for parametrizing class IV charge models is called Charge Model 2 (CM2). The CM2 mapping of BPW91/MIDI! Löwdin charges was discussed in some detail. Our investigation shows that the scheme used in CM2 is more efficient and universal than CM1. Despite the simplicity of the CM2 mapping, the CM2-charge-derived dipole moments for 36 representative molecules are found to be more accurate than the results obtained from high-level theoretical calculations based on the MP2/cc-pVDZ method and the dipole moment operator applied to the continuous electron density. Partial charges calculated from CM2 agree very well with MP2/cc-pVDZ//HF/MIDI! ChElPG charges for 22 ions. With respect to the MIDI! basis set, one can used either 5D or 6D options for *d* shells, and our results show that the CM2 mapping works equally well.

The very unique advantage of class IV charge models is that the deficiencies in the wave functions from which the partial charges are calculated are greatly reduced because the semi-empirical mapping is optimized to minimize deviation from experiment. The new charge model has a wide range of potential applications, for instance in force fields[32] and solvation models.[2,4,5]

Supporting Information Available: A full set of dipole moments for all 198 primary training-set molecules, the parameters of the CM2/BPW91/MIDI!(Mulliken) charge model, the parameters of the CM1/BPW91/MIDI!(Löwdin) charge model, error statistics for CM2/BPW91/DZVP, CM2/AM1, and CM2/PM3, and the parameters and error statistics of the HF/6-31+G* model (14 pages). This information is available on the Internet. See any current masthead page for access and ordering information.

References and Notes

(1) Bachrach, S. M. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1993; Vol. 5, p 171.

(2) (a) Tomasi, J. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981; p 257. (b) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.

(3) Warshel, A.; Russell, S. T.; Churg, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *81*, 4785.

(4) (a) Hoijtink, G. J.; Boer, E. d.; Meij, P. H. v. D.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487. (b) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 393. (c) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164. (d) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(5) (a) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901(E). (b) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. (c) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057. (d) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385. (e) Hawkins, G. D.; Cramer C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824. (f) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.

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

(7) (a) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365. (b) Del Re, G. In *Quantum Science*; Calais, J.-L., Goscinski, O., Linderberg, J., Öhrn, Y., Eds.; Plenum Press: New York, 1976; p 53.

(8) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

(9) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1989**, *11*, 361.

(10) Hall, D.; Williams, D. E. *Acta Crystallogr.* **1975**, *A31*, 56.

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

(12) (a) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc*. **1985**, *107*, 3902. (b) Dewar, M. J. S.; Zoebisch, E. G. *J. Mol. Struct. (THEOCHEM)* **1988**, *180*, 1.

(13) Stewart. J. J. P. *J. Comput. Chem*. **1989**, *10*, 221.

(14) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*; Wiley: New York, 1986.

(15) (a) Parr, R. G.; Yang, W. In *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989. (b) Ziegler, T. *Chem. Rev. (Washington, D.C.)* **1991**, *91*, 651.

(16) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098. (b) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 6533.

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R. Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision D.4*, Gaussian, Inc., Pittsburgh, PA, 1995.

(18) (a) MIDI! basis set for H, C, N, O, F, P, S, and Cl: Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281. (b) MIDI! basis set for Si, Br, and I: Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acc.*, in press.

(19) (a) Dunning, T. H., Jr. *J. Chem. Phys*. **1989**, *90*, 1007. (b) Woon, E. E.; Dunning, T. H., Jr. *J. Chem. Phys*. **1993**, *98*, 1358.

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

(21) Armstrong, D. R.; Perkins, P. G.; Stewart, J. J. P. *J. Chem. Soc., Dalton Trans*. **1973**, 838.

(22) (a) Stark, B.; In *Molecular Constants from Microwave Spectroscopy*, Landolt-Börnstein, New Series, Group II; Hellwege, K. H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1967; Vol. 4, p 136. (b) Demaison, J.;

Hütner, W.; Stark, B.; Buck, I.; Tischer, R.; Winnewisser, M. In *Molecular Constants*, Landolt-Börnstein, New Series, Group II; Hellwege, K. H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1974; Vol. 6, p 261. (c) Demaison, J.; Hütner, W.; Tiemann, E. In *Molecular Constants*, Landolt-Börnstein, New Series, Group II; Hellwege, K. H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1982; Vol. 14a, p 584. (d) Nelson, R. D.; Lide, D. R.; Maryott, A. A. *Natl. Stand.. Ref. Data Ser*. United States National Bureau of Standards, **1967**, *NSRDS.-NBS 10*.

(23) (a) Hocking, W. H. *Z. Naturforsch* **1976**, *31A*, 1113. (b) Caminati, W. *J. Mol. Spectrosc*. **1981**, *86*, 193. (c) Caminati, W.; Corbelli, G. *J. Mol. Spectrosc*. **1981**, *90*, 572.

(24) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785. (b) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. (c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem*. **1994**, *98*, 11623.

(25) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem*. **1992**, *70*, 560.

(26) Andzelm, J.; Wimmer, E. *J. Chem. Phys*. **1992**, *96*, 1280.

(27) Hawkins, G. D.; Giesen, D. J.; Chambers, C. C.; Lynch, G. C.; Rossi, I.; Storer, J. W.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. AMSOL-version 6.1, Oxford Molecular Group, 1997.

(28) Carroll, D. L. *AIAA J*. **1996**, *34*, 338.

(29) The FORTRAN source code for the genetic algorithm can be obtained from D. L. Carroll's World Wide Web site at *http://www.staff.uiuc.edu/~carroll/ga.html*.

(30) Cole, A. R. H.; Michell, A. J. *Spectrochim. Acta*, **1964**, *20*, 739.

(31) (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys*. **1980**, *72*, 650. (b) Clark, T.; Chandrasekhar, J.; Schleyer, P. v. R. *J. Comput. Chem*. **1983**, *4*, 294. (c) Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett*. **1992**, *197*, 499.

(32) (a) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Compt. Chem*. **1986**, *7*, 230. (b) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc*., **1988**, *110*, 1657. (c) Jorgensen, W. L.; McDonald, N. A.; Selmi, M.; Rablen, P. R. *J. Am. Chem. Soc*. **1995**, *117*, 11809. (d) Kaminski, G.; Jorgensen, W. L. *J. Phys. Chem*. **1996**, *100*, 18010.