

Rules for minimal atomic multipole expansion of molecular fields

E. V. Tsiper^{a)}

*Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854;
Center for Computational Material Science, Naval Research Laboratory, Washington, DC 20375;
and School of Computational Sciences, George Mason University, Fairfax, Virginia 22030*

K. Burke

Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854

(Received 24 October 2003; accepted 21 November 2003)

A nonempirical minimal atomic multipole expansion (MAME) defines atomic charges or higher multipoles that reproduce electrostatic potential outside molecules. MAME eliminates problems associated with redundancy and with statistical sampling, and produces atomic multipoles in line with chemical intuition. © 2004 American Institute of Physics. [DOI: 10.1063/1.1640995]

The problem of representing the electrostatic potential outside a molecule using atomic charges or higher atomic multipoles is very important for understanding intermolecular forces. Atomic partial charges, an important part of chemical intuition, are defined in many different ways for different purposes. Chemically-derived (CD) charges, such as Mulliken¹ or Löwdin,² often describe molecular fields poorly.³ More recent schemes partition molecular density into atomic regions, which may or may not overlap.⁴ Similar approaches have been developed for solids.⁵ Most attractive for our purposes are potential-derived (PD) charges, which avoid representation of the density by producing the “best” fit to the molecular potential directly.^{6,7} Atomic dipoles and quadrupoles⁷ are often used to increase accuracy in solvation problems⁸ and force field calculations.⁹ Induced atomic dipoles appear naturally in electronic polarization of molecular solids¹⁰ to account for the small part of molecular polarization that is due to the deformation of atomic orbitals and is not captured by redistribution of charges.

Computational schemes for PD multipoles such as Merz–Kollman (MK),¹¹ CHelp,¹² or CHelpG¹³ differ mainly in the sampling domain and the resulting atomic charges are strongly method-dependent.¹⁴ Worse still, PD methods often yield atomic charges that are counter-intuitive, such as negative charges on hydrogens in alkanes.¹⁵ Higher multipoles only increase the redundancies inherent in distributed multipole analysis, improving on the accuracy of the field at the expense of instability in the multipole values. The severity of the problem can be somewhat reduced with SVD techniques,^{14,16} or by introducing restraints.¹⁷ Our approach does not use sampling and eliminates redundancies before they appear.

We approximate the true molecular potential, $\phi(\mathbf{r})$, as a sum of multipoles of strength q_k centered at nuclear positions \mathbf{r}_i ,

$$\phi(\mathbf{r}) \approx \phi_{\text{approx}}(\mathbf{r}) = \sum_i \sum_k q_k \phi_k(\mathbf{r} - \mathbf{r}_i), \quad (1)$$

where $\phi_k(\mathbf{r})$ is the potential due to the k th multipole of unit strength: $\phi_k(\mathbf{r}) = 1/r$ for charge, $(\mathbf{n} \cdot \mathbf{r})/r^3$ for a dipole in the

direction \mathbf{n} , and so on. Since $\nabla^2 \phi_{\text{approx}} = 0$ everywhere except at \mathbf{r}_i , but $\nabla^2 \phi = 4\pi\rho(\mathbf{r})$, the atomic multipole expansion can *only* be accurate in regions where $\rho(\mathbf{r}) \approx 0$. Furthermore, ϕ on any closed surface S on which $\rho = 0$, determines $\phi(\mathbf{r})$ everywhere outside S . We therefore choose S to be an isodensity surface, $\rho(\mathbf{r}) = f$, where f is sufficiently small to ensure negligible charge beyond S , but with sufficient potential on S for a determining fit (Fig. 1). We chose ϕ_{approx} to minimize

$$\sigma^2 = S^{-1} \oint_S dS [\phi_{\text{approx}}(\mathbf{r}) - \phi(\mathbf{r})]^2 \quad (2)$$

over S which leads to a system of linear equations $\sum_k C_{mk} q_k = b_m$, where

$$C_{mk} = S^{-1} \oint_S dS \phi_m(\mathbf{r}) \phi_k(\mathbf{r}), \quad (3a)$$

$$b_m = S^{-1} \oint_S dS \phi_m(\mathbf{r}) \phi(\mathbf{r}). \quad (3b)$$

Atomic multipoles defined in this way are fully rotationally invariant, which is an issue with some PD schemes.¹³ The error σ can be compared to $\bar{\phi}$,

$$\bar{\phi}^2 = S^{-1} \oint_S dS \phi^2(\mathbf{r}). \quad (4)$$

The crucial issue remaining is the choice of a set of multipoles. We choose a minimal set, usually one scalar value per atom, and add additional multipoles to describe lone pairs when necessary, based on the Lewis structure. This carefully chosen minimal atomic multipole expansion (MAME) set avoids redundancies but is within ~ 1 mH everywhere beyond S .

We illustrate MAME with three molecules: n -pentane, which is a classic example of difficulties encountered in PD schemes; glycine (standard and zwitterion), as a typical application in biochemistry; and water, to see how general-MAME rules apply to a small polar molecule. All densities and potentials are produced on a cubic mesh by the GAUSSIAN 98 program,¹⁸ at the B3LYP/aug-cc-pVTZ level

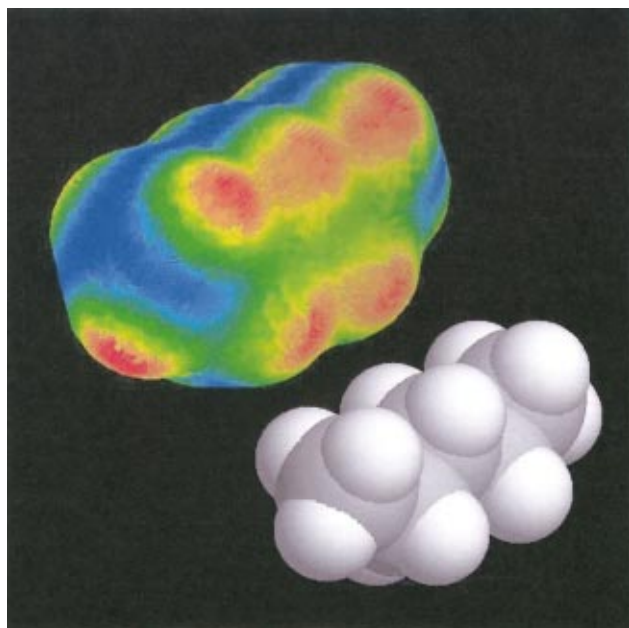


FIG. 1. (Color) Electrostatic potential over isodensity surface S of pentane. Here $f=5 \times 10^{-4}$ a.u. produces S at ~ 1.4 Å from the hydrogens and leaves $-0.2e$ charge outside. Missing charge is negligible for $f=10^{-4}$, with S at 1.8 Å.

(6-311++G** for pentane). Surface integrals (3) are computed by triangulation of S . The program runs within a few seconds, and is available on request.

Figure 1 shows $\phi(\mathbf{r})$ on S for n -pentane. Red spots ($\phi > 0$) show an excess of positive charge near each hydrogen, but all PD schemes tested yield some or all hydrogens negative. Closer inspection of Fig. 1 reveals that the positive regions occupy less solid angle around hydrogens than would be produced by a positive charge. Such a potential is consistent with a *dipole* with a negative charge pointing inwards.

Our first rule is, therefore, to assign a charge to all nuclei but protons, to which we assign a dipole moment instead. The hydrogen atom is special as its sole electron participates in the bond, leaving no electron density centered on the proton. This unique property of hydrogens is well-known in x-ray structure analysis, which systematically underestimates the C–H bond lengths for this reason.

Mulliken charges are intuitively meaningful but produce large errors in the potential (Table I). PD charges are negative on some hydrogens and still give significant errors. Adding dipoles reduces the potential error, but at the cost of producing meaningless multipoles.¹⁵ Our scheme with charges on all atoms produces similar (though better) results, but we do far better (line 2) when the charges on hydrogens are replaced with dipoles. All dipoles come out similar in magnitude (numbers in brackets, in a.u.) and point toward C within 20° of the H–C bond. The hydrogen dipoles can be safely restricted to lie along the H–C bonds (last line) with the accuracy still better than that of charges. All multipoles have reasonable values, including small charges on carbons. Note that we have now described the field outside the molecule more accurately than any existing scheme, with only

TABLE I. Partial atomic charges in n -pentane. $f=5 \times 10^{-4}$ a.u., error σ as in (2), (%) = $(\sigma/\bar{\phi})$, $e\bar{\phi}=3.6$ mHartree (1 mH=27 meV \sim kT at 300 K). “ μ ” indicates atomic dipoles, “ μ_r ”—dipoles restricted along H–C bonds.

Method	q(H), range	q(C), range	$e\sigma$, mH (%)
CD charges			
Mulliken	+0.11...+0.14	-0.59...-0.11	9.3 (260)
ZINDO	+0.03...+0.04	-0.15...-0.04	3.5 (99)
PD charges			
CHelp	-0.04...+0.04	-0.11...+0.15	3.5 (97)
CHelpG	-0.04...+0.04	-0.16...+0.16	3.2 (87)
MK	-0.03...+0.06	-0.22...+0.13	3.1 (86)
PD charges plus dipoles			
CHelp+ μ	-0.76...+0.10	-0.66...+2.05	2.8 (78)
CHelpG+ μ	-0.32...-0.30	+0.65...+0.86	1.8 (51)
MK+ μ	-0.27...-0.20	+0.50...+0.64	1.8 (49)
MAME			
Charges	-0.01...+0.09	-0.34...+0.13	2.6 (72)
μ (H)	($\mu=0.07...0.09$)	-0.01...+0.03	0.5 (15)
μ_r (H)	($\mu=0.06...0.09$)	-0.02...+0.01	1.6 (45)

one parameter per nucleus (a charge on each carbon and a bond-directed dipole on each hydrogen).

The same choice of multipoles yields a 1.05 mH error (=2%) in the glycine zwitterion, $(\text{NH}_3)^+ - \text{CH}_2 - \text{COO}^-$, down from 4% with charges alone and 4%–6% with standard PD schemes. The glycine zwitterion is highly polar with dipole $\mu=10.3$ D, which MAME recovers within 0.1% accuracy.

Table II lists MAME results for glycine in its standard form, $\text{NH}_2 - \text{CH}_2 - \text{COOH}$, and illustrates the need for special treatment of lone pairs. In the zwitterion, the NH_3 group is well-described by a charge on N and three dipoles on hydrogens, similar to methyls in pentane. The NH_2 group in glycine lacks one site, but has extra electron density associated with the lone pair. We thus assign a dipole moment to N, in addition to its charge, restricted along the sp^3 direction of the lone pair.

Similarly, each oxygen has two lone pairs. Two dipoles for the two lone pairs sum to just one dipole along the symmetry axis, leading to only one variational parameter. The potential of this single dipole, however, is axially symmetric, whereas the potential around the oxygen deviates from axial symmetry due to the particular orientation of the lone pairs. Such a deviation can be accounted for with a quadrupole moment on the oxygen. The finite system of charges sketched in the inset in Fig. 2 shows what is needed. Computing a multipole expansion of three charges we describe

TABLE II. MAME for glycine without and with lone pair multipoles. $f=10^{-4}$ a.u., $e\bar{\phi}=15$ mH.

	NH ₂	CH ₂	C=	=O	-OH	$e\sigma$, mH (%)
μ_r (H)	-0.05	-0.03	+0.79	-0.54	-0.17	4.1 (27)
μ_r (H)	+0.06	+0.11	+0.37	-0.67	+0.13	
+ μ_r (N,O)	-0.65			+0.19	-0.71	1.6 (11)
+ θ_r (O)				-0.53	-0.87	

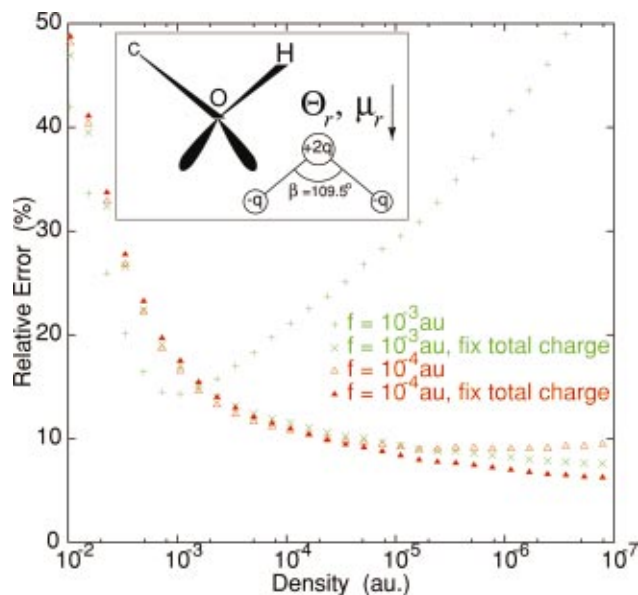


FIG. 2. (Color) MAME accuracy away from glycine. Ratio of $(\phi_{\text{approx}} - \phi)$ to ϕ , both square-averaged over points with a given density ρ , which is an inverse measure of distance. The inset illustrates two lone pairs on oxygen represented with a combination of restricted dipole μ_r and quadrupole θ_r .

two lone pairs with two scalar parameters, a dipole μ_r restricted along the symmetry axis, and a *restricted quadrupole* θ_r which has angle β as a fixed parameter ($\beta=120^\circ$ for O= and 109.7° for O-, due to sp^2 and sp^3 hybridization respectively). θ_r and μ_r are chosen negative with “-” pointing outside. Table 2 shows a clear advantage of such a multipole set.

To make the definition more transparent, θ_r can be expressed in the conventional form as a carefully crafted combination of $\theta_{zz} = \theta_r(3 \cos^2 \beta/2 - 1)$, $\theta_{yy} = \theta_r(3 \sin^2 \beta/2 - 1)$, $\theta_{xx} = -\theta_r$, and $\theta_{xy} = \theta_{xz} = \theta_{yz} = 0$, which depend on a single scalar parameter θ_r . Here z is along the symmetry axis, and the lone pairs are in the yz -plane. A single restricted quadrupole of strength θ_r creates the potential

$$\phi_{quad}(\mathbf{r}) = \theta_r \frac{3(\mathbf{r} \cdot \mathbf{n}_1)^2 + 3(\mathbf{r} \cdot \mathbf{n}_2)^2 - 2r^2}{2r^5}, \quad (5)$$

where \mathbf{n}_1 and \mathbf{n}_2 are the directions of the lone pairs.

For the zwitterion, MAME does not require θ_r on the oxygens, because of the resonance. Lone pairs in the sp^2 and sp^3 configurations lie in perpendicular planes, virtually destroying any asymmetry.

MAME accuracy improves away from the molecule (Fig. 2). If f is too large (10^{-3}), there is a net charge inside S which strongly affects the asymptotic behavior. This can be repaired by fixing the total charge using a Lagrange multiplier.¹⁶ Figure 2 demonstrates MAME’s insensitivity to choice of f , provided the total charge is correct.

Last, we analyze the water molecule. The first rule leads to two dipoles on hydrogens pointing along the bonds, sketched in Fig. 3. The charge on oxygen is zero because the molecule is neutral. The dipoles are equal due to symmetry and require no calculation since their vector sum must yield

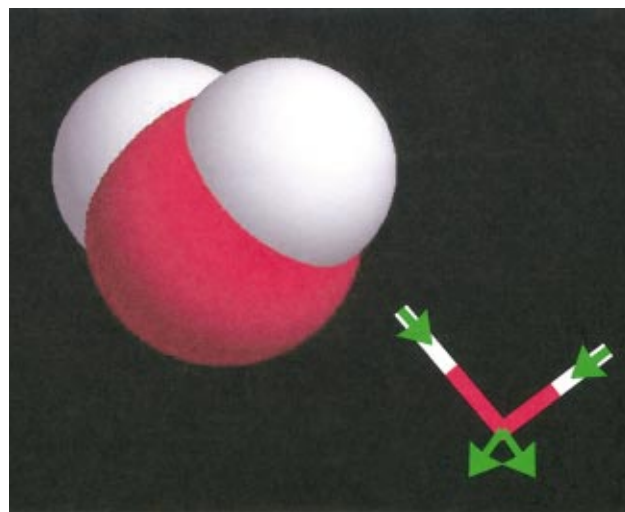


FIG. 3. (Color) Schematic MAME representation of water.

the dipole moment of water, 1.847D (B3LYP/aug-cc-pVTZ value). This already reduces the error to 21%, from 45% with a single dipole on oxygen. The two distributed dipoles yield $\Theta_{xx} - \Theta_{yy} = 4.06 \text{ D}\text{\AA}$ for the quadrupole moment of water, whereas experiment¹⁹ gives 5.12 D\AA. We note that $\Theta_{xx} - \Theta_{yy}$ is the only invariant combination of quadrupole components, since the finite dipole makes them dependent on the center of coordinates.

We next add μ_r and θ_r multipoles on the oxygen to describe the correction due to lone pairs, and we find excellent accuracy $e\sigma = 0.59 \text{ mH}$ ($<3\%$) on and beyond S (1.56 \AA from H and 2.11 \AA from O, $f=10^{-4}$). The sp^3 choice of $\beta=109.5^\circ$ in θ_r is crucial: accuracy deteriorates dramatically (to 12%) when β is changed to, e.g., 180° (θ_r replaced with θ_{yy}). The sp^3 description of the oxygen lone pairs is appropriate due to invariance under unitary rotations of occupied orbitals.

In conclusion, molecular fields are represented to chemical accuracy with a minimal set of atomic multipoles carefully chosen based on the Lewis structure of the molecule. All H atoms are represented as dipoles. Lone pairs are treated with extra multipoles, avoiding additional off-nuclear expansion sites.²⁰ The scheme yields multipole values that conform to chemical intuition, are unique, fully rotationally-invariant and free of sampling errors.

We thank K. Krogh-Jespersen, Z. G. Soos, M. Lobanov, and R. Gaudoin for enlightening discussions. E. T. is grateful to T. M. Burkat for encouragement. This work was funded by AFOSR and the New Jersey Commission on Science and Technology.

^aElectronic mail: tsiper@dave.nrl.navy.mil

¹R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

²P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).

³S. M. Bachrach, Rev. Comput. Chem. **5**, 171 (1994).

⁴P. Politzer and R. R. Harris, JACS **92**, 6451 (1970); A. J. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985); **56**, 1065 (1985); R. F. W. Bader *Atoms in Molecules—A Quantum Theory* (Oxford University Press, Oxford, 1990); R. F. Nalewajski and R. G. Parr, J. Phys. Chem. A **105**, 7391 (2001); C. E. Whitehead, C. M. Breneman, N. Sukumar, and M. D. Ryan, J. Comput. Chem. **24**, 512 (2003).

- ⁵L. L. Boyer, H. T. Stokes, and M. J. Mehl, *Phys. Rev. Lett.* **84**, 709 (2000).
⁶F. A. Momany, *J. Phys. Chem.* **82**, 592 (1978).
⁷D. E. Williams, *J. Comput. Chem.* **9**, 745 (1988).
⁸G. J. Tawa, R. L. Martin, L. R. Pratt, and T. V. Russo, *J. Phys. Chem.* **100**, 1515 (1996).
⁹P. Ren and J. W. Ponder, *J. Comput. Chem.* **23**, 1497 (2002).
¹⁰E. V. Tsiper and Z. G. Soos, *Phys. Rev. B* **64**, 195124 (2001); Z. G. Soos, E. V. Tsiper, W. Gao, and A. Kahn, *Chem. Phys. Lett.* **360**, 47 (2002); E. V. Tsiper and Z. G. Soos, *Phys. Rev. B* **68**, 085301 (2003).
¹¹B. H. Besler, K. M. Merz, and P. A. Kollman, *J. Comput. Chem.* **11**, 431 (1990).
¹²L. E. Chirlian and M. M. Francl, *J. Comput. Chem.* **8**, 894 (1987).
¹³C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.* **11**, 361 (1990).
¹⁴E. Sigfridsson and U. Ryde, *J. Comput. Chem.* **19**, 377 (1998).
¹⁵D. E. Williams, *J. Comput. Chem.* **15**, 719 (1994).
¹⁶M. M. Francl, C. Carey, L. E. Chirlian, and D. M. Gange, *J. Comput. Chem.* **17**, 367 (1996); K. Hinsien and B. Roux, *ibid.* **18**, 368 (1997).
¹⁷C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**, 10269 (1993).
¹⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.11.1, Gaussian, Inc., Pittsburgh, PA, 2001.
¹⁹J. Verhoeven and A. Dymanus, *J. Chem. Phys.* **52**, 3222 (1970).
²⁰R. W. Dixon and P. A. Kollman, *J. Comput. Chem.* **18**, 1632 (1997).