Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model

S. R. Cox and D. E. Williams

Department of Chemistry, University of Louisville, Louisville, Kentucky 40292
Received 1 August 1980; accepted 16 January 1981

Electrostatic potentials and Mulliken net atomic charges were calculated from STO-3G, 6-31G, and 6-31G** SCF-MO wavefunctions for hydrogen fluoride, water, ammonia, methane, acetylene, ethylene, carbon dioxide, formaldehyde, methanol, formamide, formic acid, acetonitrile, diborane, and carbonate ion. In each case optimized net atomic charges (potential-derived charges) were also obtained by fitting the electrostatic potentials calculated directly from the wavefunctions in a shell enveloping the molecules outside of their van der Waals surfaces. The electrostatic potentials calculated from the potential-derived charge distributions were then compared with the defined quantum mechanical electrostatic potentials and with the electrostatic potentials of the Mulliken charge distributions.

INTRODUCTION

The concept of net atomic charges in molecules is a useful one which permeates nearly every facet of the study of molecules. Such net atomic charges have been used as an index of reactivity of atomic sites in molecules, as an aid to understanding the molecular dipole moment, for interaction energy calculations in molecular complexes, for packing energy calculations in crystals, and for a variety of other purposes. Unfortunately, however, the net atomic charge is not an exactly defined physical property.

In the interior of the molecule near the nuclei a large positive electrostatic potential exists. To the crudest approximation, outside the nominal boundary of the electron cloud of a neutral molecule, the electrostatic potential is zero. However, it has been long known since the earliest studies on electronegativity of atoms that the electron cloud of heteroatomic molecules does not usually distribute itself in such a way as to give a zero electropotential. Quite to the contrary, a significant electrostatic potential exists around even a neutral molecule, and this potential is of utmost importance in defining the reactivity and properties of the molecule.

A recent review by Scrocco and Tomasi[1] discusses the calculation of the molecular electrostatic potential from the molecular wavefunction and the relationship of this potential to molecular reactivity and intermolecular forces. Unlike net atomic charges, the molecular electrostatic potential is a rigorously defined quantum mechanical property. Furthermore, since the electrostatic potential is the expectation value of a one-electron operator, the SCF-MO calculation of the electrostatic potential is correct to one order higher than the wavefunction used.

The recent rapid development of *ab initio* self-consistent field molecular orbital (SCF-MO) calculations, made possible by the use of electronic computers, has made available rather accurate wavefunctions for smaller molecules, up to those containing 50 electrons or so. We can expect a further extension of the SCF-MO method to larger molecules in the near future. Nevertheless, it seems unlikely that the *ab initio* SCF-MO method will be extended to very large molecules such as proteins or nucleotides, without using some approximations. The net atomic charge concept is useful for evaluating the interaction energy between such large molecules, and in many cases for evaluating the interaction energies involving smaller molecules as well.

Until quite recently virtually the only method in common use for representation of net atomic

Journal of Computational Chemistry, Vol. 2, No. 3, 304–323 (1981)

CCC 0192-8651/81/030304-20$04.00

charges in polyatomic molecules was the population analysis method of Mulliken.[2] The Mulliken charges are not defined physical observables in terms of quantum mechanics (QM). Their widespread use is a consequence of the fact that they do show trends which appear chemically realistic and compatible with concepts of electronegativity of atoms in molecules. We will show that for the molecules considered the Mulliken charges do not give a very accurate portrayal of the molecular electrostatic potentials.

Momany[3] determined net atomic charges by empirically fitting the SCF-MO calculated electrostatic potentials (QM electrostatic potentials) of formamide, methanol, and formic acid. This article refines and extends the work of Momany. We have selected a set of simple molecules for which we calculate the Mulliken charges (M charges) and also fit net atomic charges to the QM electrostatic potentials. We make detailed numerical and pictorial comparisons between the M charges and QM electrostatic potential-derived net atomic charges (PD charges) for SCF-MO wavefunctions of varying complexity.

CALCULATION OF THE MOLECULAR ELECTROSTATIC POTENTIAL

The electrostatic potential for a unit positive charge at point $\mathbf{r}$ in the vicinity of the given molecule is calculated in the following way:

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \sum_{mn} P_{mn} \int \frac{\phi_m \phi_n}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'$$

P_{mn} refers to a particular density matrix element of the wavefunction for the molecule as obtained from the SCF-MO calculation; ϕ_m are the atomic basis wavefunctions. The first term is summed over all nuclei, and the second term is summed over all electrons in the molecule. In this study, three atomic wavefunction basis sets[4,5] of the Gaussian type were used: STO-3G, 6-31G, and 6-31G**. The molecular electrostatic potential was calculated with each basis set for each molecule, and the Mulliken charges were calculated in each case for the purpose of comparison with the PD charges. The program[6] GAUSSIAN 79 was used with a DEC-10 computer. The average relative central processor time required for calculations in the STO-3G, 6-31G, and 6-31G** basis sets was 1.0 to 3.0 to 7.5. The largest 6-31G** calculation ($CHONH_2$) required 190 min of central processor time.

SELECTION OF POTENTIAL SAMPLING POINTS AND FITTING PROCEDURE

It was necessary to establish criteria for the selection of points where the molecular electrostatic potential was to be evaluated and subsequently fitted by a net atomic charge model. The region we are interested in is a shell that is just outside the van der Waals radii of the atoms in the molecule. The electrostatic potential in this region (as in other regions) is an exactly defined physical property in quantum mechanics. The numerical value of the electropotential at any point may be found using the electrostatic potential operator and the wavefunction of the molecule.

As mentioned above, it is necessary to avoid getting too close to nuclei, where the electrostatic potential is always positive. Well-established values of van der Waals radii are available, which essentially define empirically the boundary of significant electron density. To include the possibility of interactions of the reference molecule with heteroatoms, we choose the van der Waals radius of hydrogen, the smallest atom considered here, as a possible atom in an interacting molecule.

Thus we define the molecular boundary as the van der Waals surface, and extend this boundary 1.2 Å farther to allow for the possibility of an interaction with hydrogen, the smallest atom considered. Using the formic acid STO-3G calculation as an example, we tested the effect of choosing the molecular boundary smaller or larger by 1 Å. The main effect observed was that the fit to the QM potential (both in the rms fit and in the relative rms fit, defined below) became worse at the smaller boundary and better at the larger boundary. The PD charges themselves changed by a negligible amount.

We also investigated the effect of varying the shell thickness t and the cubic grid spacing used, d. As t increases and d decreases the molecular

Table I. Mulliken (M) and potential-derived (PD) net atomic charges ($\times 10^3$) in electron units, for various molecules using three Gaussian basis sets for the SCF calculation. The rms fits σ (in kJ/mol) to the QM electrostatic potentials by the M and the PD model electrostatic potentials are also given for each molecule, as well as the relative percentage rms fits s. The standard deviations for the individual PD charges are given in parentheses; these reflect only the goodness of fit to the QM electrostatic potential and do not reflect any errors in the values calculated for the potential itself.

		STO-3G		6-31G		6-31G**	
Molecule	Atom	M	PD	M	PD	M	PD
HF	F	−211	−294(1)	−482	−524(1)	−395	−449(3)
	H	211	294(1)	482	524(1)	395	449(3)
	σ	6.1	0.5	3.1	0.6	4.4	1.6
	s	28.5	2.2	8.3	1.6	14.1	5.0
H_2O	O	−366	−614(1)	−792	−940(3)	−674	−786(6)
	H	183	307(1)	396	470(2)	337	393(3)
	σ	11.3	0.5	5.5	1.2	5.5	2.1
	s	40.4	1.6	16.4	3.0	16.6	6.1
NH_3	N	−468	−972(4)	−909	−1278(8)	−789	−1047(9)
	H	156	324(1)	303	426(3)	263	349(3)
	σ	13.2	0.8	9.7	1.6	7.4	2.0
	s	52.0	3.0	29.8	5.0	26.6	7.3
CH_4	C	−252	−484(4)	−620	−528(6)	−472	−556(6)
	H	63	121(1)	155	132(2)	118	139(1)
	σ	0.8	0.2	0.8	0.4	0.7	0.3
	s	49.3	7.8	22.0	9.5	20.1	9.4
C_2H_2	C	−108	−180(1)	−323	−286(1)	−236	−297(1)
	H	108	180(1)	323	286(1)	236	297(1)
	σ	3.8	0.1	1.7	0.1	3.3	0.2
	s	39.7	1.1	12.8	0.6	20.8	1.0
C_2H_4	C	−124	−164(4)	−328	−356S5)	−256	−352(4)
	H	62	82(2)	164	178(3)	128	176(2)
	σ	1.4	0.7	1.4	0.9	2.6	0.8
	s	34.6	18.1	16.4	10.3	29.9	9.5
CO_2	O	−232	−405(3)	−432	−569(3)	−453	−434(2)
	C	464	810(5)	864	1138(6)	906	868(4)
	σ	5.9	0.1	3.9	0.4	1.0	0.5
	s	42.6	0.6	18.7	2.2	6.9	3.8
CH_2O	O	−192	−293(7)	−450	−522(6)	−443	−459(4)
	C	83	359(2)	154	530(18)	242	423(13)
	H	55	−33(7)	148	−4(6)	100	18(4)
	σ	3.6	1.6	6.9	1.3	4.7	0.9
	s	19.1	8.5	19.3	3.7	15.6	2.9
CH_3OH	O	−307	−501(13)	−736	−827(15)	−646	−680(13)
	C	−62	188(13)	−129	416(60)	−14	244(52)
Eclipsed	H	72	14(4)	180	14(16)	128	48(14)
Staggered	H	53	−9(14)	143	−44(16)	97	−18(14)
Hydroxyl	H	191	317(5)	399	485(6)	338	424(5)
	σ	7.1	1.6	12.2	1.9	11.9	1.6
	s	36.0	8.1	39.2	6.0	46.0	6.2
$HCONH_2$	O	−272	−435(9)	−546	−666(9)	−550	−556(9)
	N	−441	−841(2)	−912	−1114(26)	−748	−883(24)
	C	265	676(3)	543	890(29)	567	637(27)
Carbon	H	52	−62(9)	164	−15(10)	105	30(9)
Trans	H	199	323(8)	371	435(9)	308	374(8)
Cis	H	197	337(8)	380	470(8)	318	398(8)
	σ	6.5	1.6	10.3	1.7	11.8	2.2
	s	24.0	5.7	23.3	3.9	20.4	3.8
HCOOH	O	−298	−554(8)	−678	−817(8)	−583	−628(6)
Carbonyl	O	−263	−439(7)	−517	−676(7)	−531	−568(5)
	C	268	699(17)	553	948(17)	601	674(14)
	H	70	−61(6)	209	4(7)	143	59(5)

Table I *(continued from previous page)*

Molecule	Atom	STO-3G		6-31G		6-31G**	
		M	PD	M	PD	M	PD
Hydroxyl	H	223	355(5)	433	541(4)	370	462(4)
	σ	7.8	1.4	13.4	1.4	10.9	1.2
	s	47.2	8.7	49.1	5.3	42.4	4.6
CH_3CN	N	−196	−457(3)	−329	−555(2)	−463	−514(18)
	C	−177	−574(20)	−472	−602(22)	−376	−577(17)
Nitrile	C	76	515(10)	135	557(7)	296	488(6)
	H	99	172(7)	222	200(5)	181	201(4)
	σ	8.3	0.7	5.1	0.5	4.1	0.4
	s	27.9	2.2	19.4	1.9	15.6	1.4
B_2H_6	B	86	−109(17)	89	222(25)	124	178(33)
	H	−40	−31(5)	−27	−120(7)	−51	−115(4)
Bridging	H	−6	171(8)	−35	18(11)	−22	52(15)
	σ	3.8	0.4	7.5	1.0	3.8	0.8
	s	73.7	7.2	95.2	13.0	75.4	15.0
CO_3^{2-}	O	−725	−938(1)	−934	−1152(2)	−936	−1070(4)
	C	175	814(3)	802	1456(7)	808	1210(11)
	σ	3.5	0.3	8.1	0.7	7.4	1.1
	s	1.1	0.1	1.1	0.1	0.7	0.1

electric potential must by calculated at a larger number of points. Since the electrostatic potential is largest and most detailed closer to the molecule, if we fit this shell the region outside the shell is well represented also. So it is not necessary to consider a large number of additional grid points farther away from the molecule.

We performed a number of test calculations with various values of d and t with an eye toward reducing the number of points to a minimum commensurate with good definition of the net atomic charges. We found that d values of 1.0–1.2 Å and a t value of 1 Å were satisfactory. These choices of d and t gave around 100–200 sampling points for the molecular electrostatic potential. Of course, for symmetrical molecules only the asymmetrical portion of the grid needs consideration.

The molecular electrostatic potentials at the grid points were fitted by a set of point charges located at the nuclei: This is the PD net atomic charge model for the electrostatic potential of the molecule. The criterion of fit was the sum-of-squares function defined as follows:

$$R = \sum_i^m w_i \left[V_i^0 - \sum_j^{n-1} q_j r_{ij}^{-1} + \left(\sum_j^{n-1} q_j \right) r_{in}^{-1} - Z r_{in}^{-1} \right]^2$$

where V_i^0 is the calculated QM electrostatic potential at point i, q_j is the net charge on atom j (a variable), r_{ij} is the distance between atom j and the ith grid point, m is the number of grid points, n is the number of atoms, and Z is the net charge on the molecule, which is zero for nonionic species. The last two terms of R reflect the condition that the net atomic charges must sum to Z.

To find the minimum value of R, the first partial derivatives of R with respect to each of the $(n-1)$ independent net atomic charges were obtained, and the resulting linear equations solved in the usual least-squares fashion. Appropriate row and column operations were performed to reduce the matrix equations when certain charges were assumed to be equal by symmetry. If s is the number of symmetry conditions, the resulting matrix equations are of order $(n-s-1)$. The statistical weight w_i for each point was taken as unity. This choice of weights was found satisfactory for the relatively thin (1 Å) shell.

RESULTS AND COMPARISONS OF THE NET ATOMIC CHARGES

M net atomic charges and PD net atomic charges for a variety of small molecules are given in Table I. The results with three different atomic basis sets are shown: the minimal STO-3G basis, the split-

Table II. Least-squares scale factors and rms fits of model net atomic charges to 6-31G** PD charges as reference values.

Model	Scale factor	rms fit (e)
6-31G** PD	(1.00)	(0.000)
6-31G PD	0.82	0.049
STO-3G PD	1.13	0.078
6-31G** M	1.14	0.086
STO-3G M	2.16	0.113
6-31G M	1.02	0.121
DZ M	0.63	0.272

valence 6-31G basis, and the extended 6-31G** basis. The computer time required to obtain the SCF-MOs increases in going from STO-3G to 6-31G to 6-31G** basis sets. In order to make a quantitative comparison between the model electrostatic potentials, the table shows the rms fits

$$\sigma = \left[N^{-1} \sum_i (V_i^0 - V_i^{\text{model}})^2\right]^{1/2}$$

and the relative rms fits

$$s = \sigma \Big/ \left[N^{-1} \sum_i (V_i^0)^2\right]^{1/2}$$

of the M and PD model electrostatic potentials to the QM electrostatic potentials.

Table I shows that the magnitudes of the STO-3G M charges for the first 12 molecules are generally smaller than the corresponding 6-31G** M charges for the same atoms and molecules. The least-squares scale factor between the two sets of charges is 1.91 and the resulting rms fit between them is 0.055e. This situation shows the sensitivity of the M charges to the choice of basis set, and also that the magnitudes of the STO-3G M charges differ mainly from the magnitudes of the 6-31G** M charges by a scaling constant. The relation between the 6-31G and 6-31G** M charges is a scale factor of 0.90, yielding an rms fit of 0.067e. Thus, the 6-31G M charges are generally larger than the 6-31G** values, and again differ mainly by a scaling constant.

A similar comparison for the PD charges shows that the STO-3G PD charges may be related to 6-31G** PD charges by a scale factor of 1.13 to give an rms fit of 0.078e. The 6-31G PD charges may be related to 6-31G** PD charges by a scale factor of 0.82 to give an rms fit of 0.049e. This behavior shows that for these molecules the PD charges do not require as much scaling as the M charges, and are thus less sensitive to the choice of basis set used for the SCF-MO calculation.

It is also true, for these molecules, that the agreement between M and PD charges becomes better for the extended 6-31G** basis set. The rms fit of STO-3G M charges to STO-3G PD charges is 0.127e, with a scale factor of 1.86; for 6-31G M charges to 6-31G PD charges, 0.175e, with a scale factor of 1.23; and for 6-31G** M charges to 6-31G** PD charges, 0.086e, with a scale factor of 1.14. Thus, while the net atomic charges obtained by the two methods are very different for the STO-3G basis set, this difference becomes much less for the extended 6-31G** basis set. We think that attaching too much significance to the improved agreement between M and PD charges obtained with the 6-31G** basis set would be premature, since Kern and Karplus,[7] for example, have shown that large, accurate (in terms of molecular energy) basis sets can give very different M charges for the hydrogen fluoride molecule.

We can further illustrate the sensitivity of the M charges to the choice of basis set with the double-zeta (DZ) basis of Snyder and Basch.[8] This basis is a split-shell one similar in concept to 6-31G, but with inner shells being split into two contractions as well as the valence shells. The rms fit between the DZ and 6-31G M charges is 0.246e and the scale factor is 0.70, indicating a poor fit between the two. We have calculated the QM electrostatic potential for formaldehyde using the DZ basis. We found that the DZ QM potential was very similar to the 6-31G QM potential; the DZ PD charges are as follows: oxygen, −0.530e; carbon, 0.520e; and hydrogen, 0.005e. These PD charges are almost the same as the 6-31G PD charges (Table I). The DZ M charges are oxygen, −0.306e; carbon, 0.016e; and hydrogen, 0.145e. These DZ M charges are very different from the 6-31G M charges.

If we take the 6-31G** PD charges as reference values, the scale factors and resulting rms fits to them from the other sets of charges are given in Table II (see the Discussion section).

Table III gives the observed and calculated di-

Table III. Dipole moments calculated from the M and PD net atomic charges (u_{emp}) for three atomic basis sets; dipole moments calculated directly from the SCF-MO wavefunctions (u_c) and the observed dipole moments of several molecules.

Molecule/ obs. dipole		STO-3G M	STO-3G PD	6-31G M	6-31G PD	6-31G** M	6-31G** PD
HF	u_{emp}	0.930	1.296	2.123	2.308	1.741	1.981
1.82[a]	u_c	1.288		2.296		1.973	
H_2O	u_{emp}	1.031	1.727	2 228	2.647	1.896	2.211
1.846[b]	u_c	1.726		2.630		2.184	
NH_3	u_{emp}	0.856	1.780	1.661	2.340	1.444	1.917
1.468[c]	u_c	1.783		2.319		1.887	
CH_2O	u_{emp}	1.424	1.510	3.446	3.006	3.134	2.765
2.339[d]	u_c	1.525		3.010		2.759	
CH_3OH	u_{emp}	1.175	1.491	3.087	2.357	2.467	1.899
1.7[e]	u_c	1.500		2.367		1.915	
$HCONH_2$	u_{emp}	2.104	2.596	3.957	4.205	3.791	3.929
3.72[f]	u_c	2.616		4.200		3.940	
HCOOH	u_{emp}	0.804	0.787	1.565	1.402	1.669	1.585
1.415[g]	u_c	0.789		1.399		1.583	
CH_3CN	u_{emp}	2.455	3.046	4.365	4.131	4.703	4.092
3.915[h]	u_c	3.063		4.132		4.092	

[a] R. Weiss, *Phys. Rev.*, **131**, 659 (1963).
[b] G. Birnbaum and S. K. Chatterjee, *J. Appl. Phys.*, **23**, 220 (1952).
[c] D. K. Coles, W. E. Good, J. K. Bragg, and A. H. Sharbaugh, *Phys. Rev.*, **82**, 877 (1951).
[d] J. N. Shoolery and A. H. Sharbaugh, *Phys. Rev.*, **82**, 95 (1951).
[e] R. D. Nelson, Jr., D. R. Lide, Jr., and A. A. Maryott, *Natl. Stand. Ref. Data Ser. 10* (1967).
[f] C. C. Costain and J. M. Dowling, *J. Chem. Phys.*, **32**, 158 (1960).
[g] H. Kim, R. Keller, and W. D. Gwinn, *J. Chem. Phys.*, **37**, 2748 (1962).
[h] P. A. Steiner and W. Gordy, *J. Mol. Spectrosc.*, **21**, 291 (1966).

pole moments, and Table IV gives quadrupole moments for nondipolar molecules. It is known that the 6-31G** basis, and especially the 6-31G basis, generally yields too large dipole and quadrupole moments, and this is apparent from an examination of the tables. Still larger basis sets and inclusion of configuration interaction will usually cause the observed dipole moments to be approached from the top side.

The dipole moments are valuable in evaluation of the models because reference observed values are available. Table V shows how each model compares with the observed dipole moments. Recall that the observed dipole moments are not being fitted, only the QM electrostatic potential was fitted. The table shows that the PD-model-distributed dipole moments are very similar to the QM-calculated dipole moments. The scale factors

Table IV. Quadrupole moments calculated from the M and PD net atomic charges (Q_{emp}) for three atomic basis sets; quadrupole moments calculated directly from the SCF-MO wavefunctions (Q_c) and the observed quadrupole moments of several molecules.

Molecule obs. quadrupole		STO-3G M	STO-3G PD	6-31G M	6-31G PD	6-31G** M	6-31G** PD
C_2H_2	Q_{emp}	2.489	4.148	7.443	6.590	5.438	6.844
3.0[a]	Q_c	4.160		6.590		6.829	
CO_2	Q_{emp}	−3.004	−5.245	−5.596	−7.371	−5.868	−5.620
4.3[a]	Q_c	−5.258		−7.307		−5.503	
C_2H_4	Q_{emp}	0.762	1.002	2.013	2.175	1.622	2.155
1.5[a,b]	Q_c	0.652		1.740		1.751	

[a] D. E. Stogryn and A. P. Stogryn, *Mol. Phys.*, **11**, 371 (1966).
[b] Experimental Q_{zz} value.

Table V. Unscaled and scaled rms fits of the calculated dipole moments to the observed dipole moments.

Model	Unscaled rms fit	Scaled	
		Scale factor	rms fit
6-31G** QM	0.28	0.91	0.16
6-31G** PD	0.29	0.91	0.16
6-31G** M	0.49	0.86	0.30
6-31G QM	0.59	0.85	0.37
6-31G PD	0.59	0.83	0.32
6-31G M	0.68	0.82	0.40
STO-3G QM	0.66	1.30	0.32
STO-3G PD	0.66	1.27	0.41
STO-3G M	1.00	1.68	0.17

for the three basis sets suggest that the reasons for over- or underestimation of the net atomic charges and dipole moments are similar and lie in the SCF-MO calculation and choice of basis set.

We suspect that the good agreement of the scaled STO-3G M-model-distributed dipole moments with the observed values is fortuitous; note that the unscaled fit is poor. For the larger basis sets, the QM and PD model dipole moments are always better than the M model dipole moments, scaled or unscaled.

The observed quadrupole moments are not known as accurately as are the dipole moments, so we take as reference values the 6-31G** QM quadrupole moments, instead of the observed values. Table IV shows that the 6-31G** PD values are rather close to the reference values. For acetylene and carbon dioxide the PD model quadrupole moments are always improved over the M model values. For ethylene the PD values for 6-31G and 6-31G** basis sets are a little too large.

COMPARISON OF THE MODEL ELECTROSTATIC POTENTIALS

The left side of Fig. 1 shows the QM electrostatic potential of hydrogen fluoride calculated from the STO-3G wavefunction. The right side of the figure shows the difference map of the M charge electrostatic potential minus the QM potential; it is seen that the M charge potential is too small in magnitude. The rms fit of the maps is 6.1 kJ/mol and the relative rms fit is 28.5%. The center of the figure, which shows the difference map of the PD

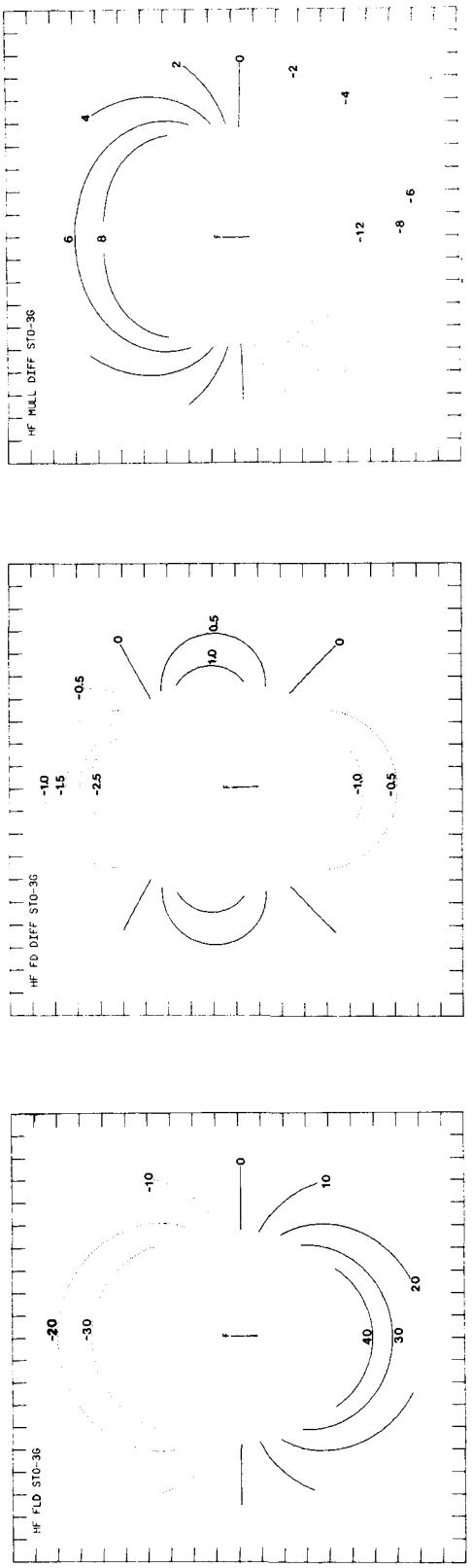

Figure 1. (*left*) The QM electrostatic potential of hydrogen fluoride calculated from the STO-3G SCF-MO wavefunction. (*center*) The electrostatic potential of the PD charges minus the QM potential. (*right*) The electrostatic potential of the M charges minus the QM potential. The molecule is drawn to scale in the center of the maps; distances are ticked in atomic units (0.52917×10^{-10} m). Contours are drawn in units of kJ/mol.

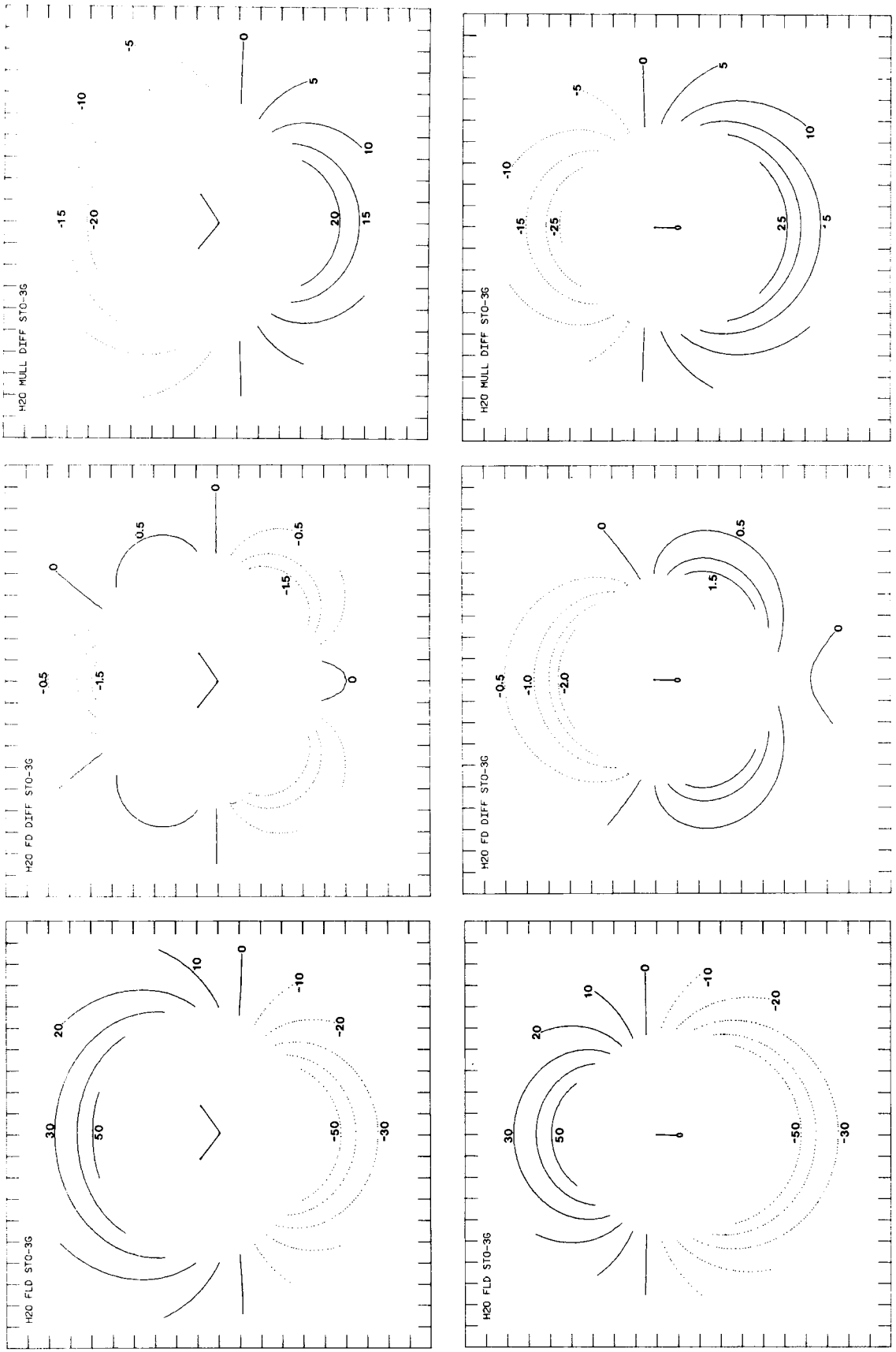

Figure 2. As in Fig. 1, for water. The bottom row shows the potential perpendicular to the plane of the molecule.

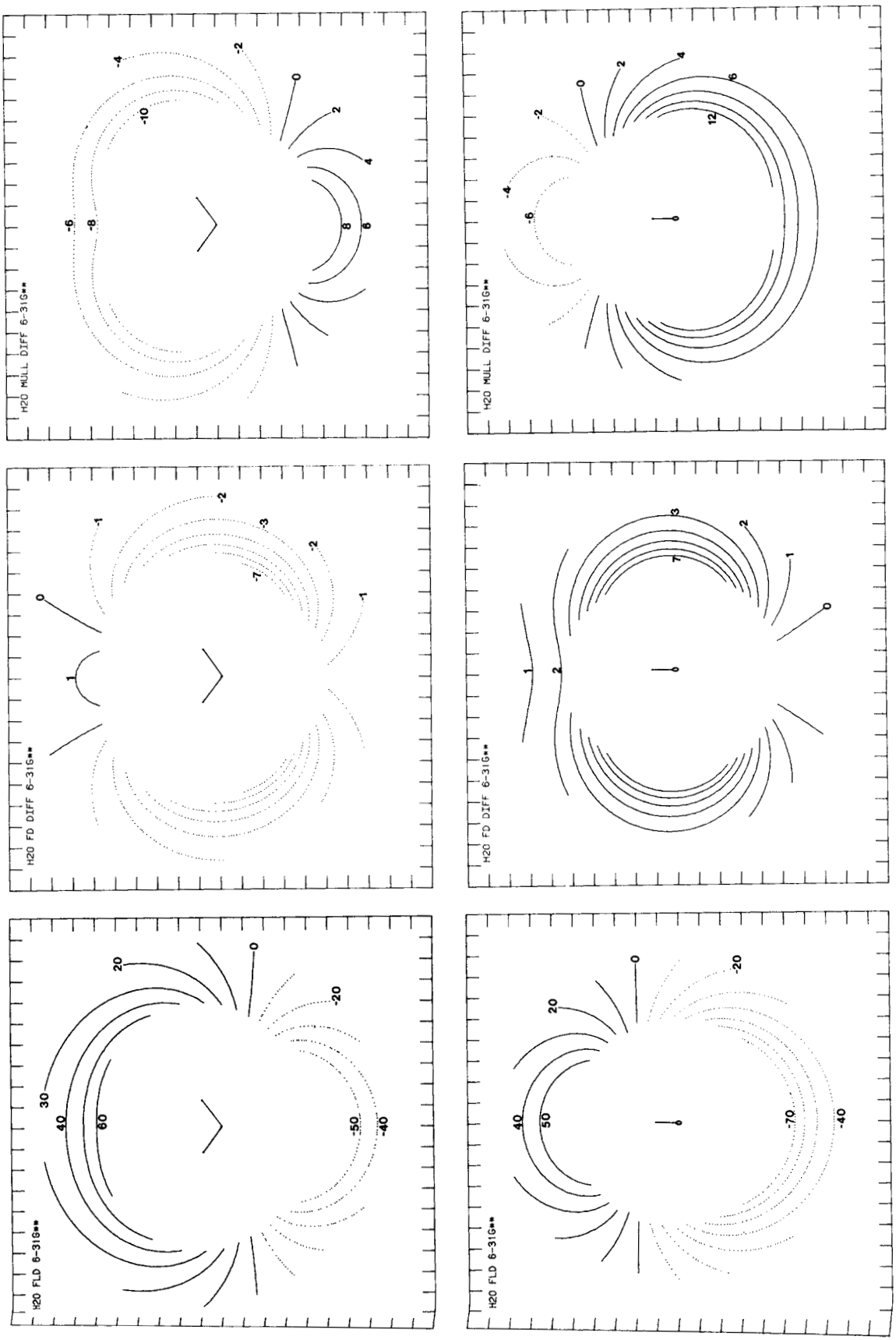

Figure 3. As in Fig. 2, but for the 6-31G** SCF-MO wavefunction.

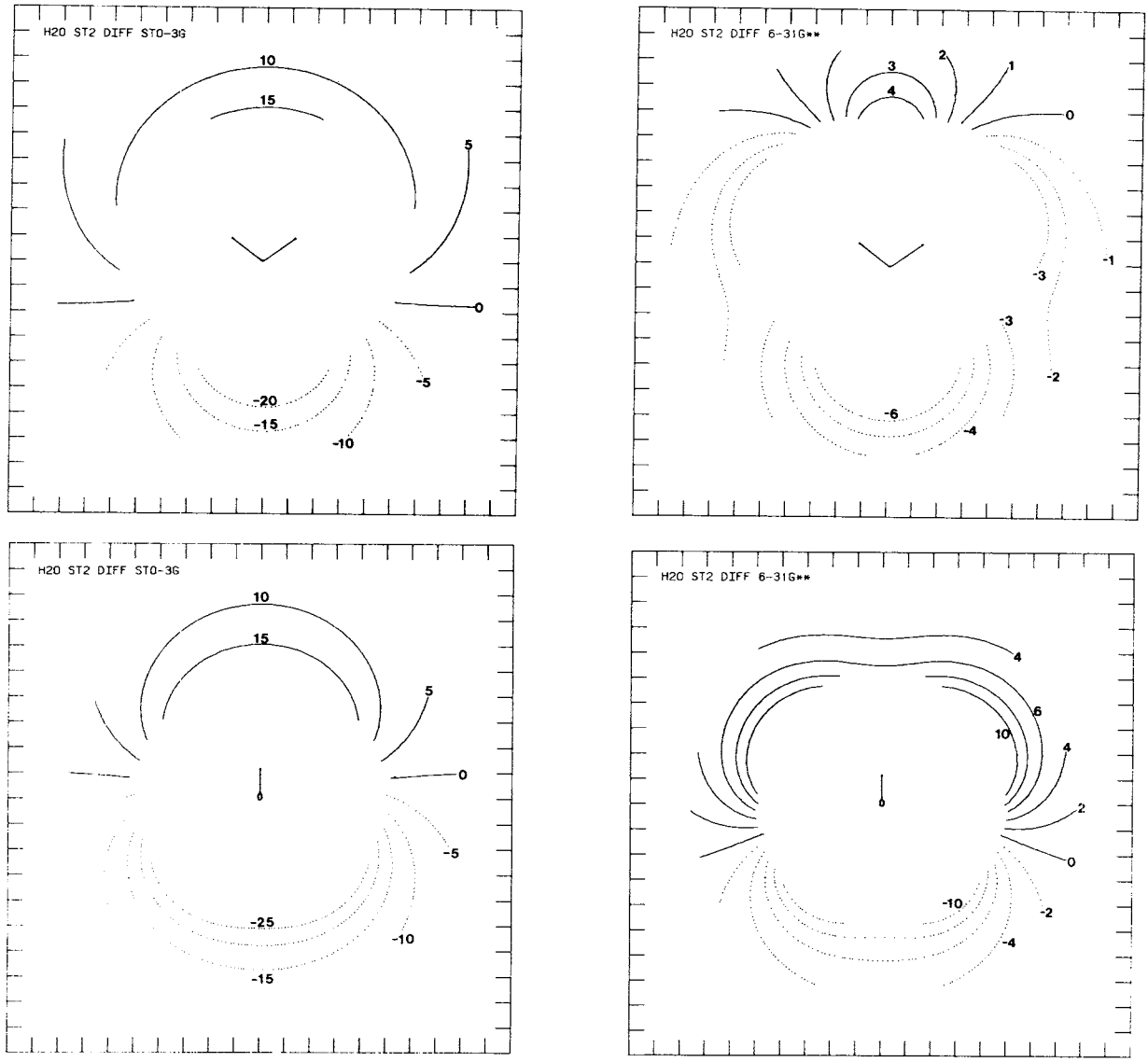

Figure 4. (*left*) The ST2 potential of water minus the QM potential calculated from the STO-3G SCF-MO wavefunction. (*right*) The ST2 potential of water minus the QM potential calculated from the 6-31G** SCF-MO wavefunction. The bottom row shows the difference potentials perpendicular to the plane of the molecule.

charge electrostatic potential minus the QM potential, indicates that the PD charge potential gives a much better fit than the M charge potential. The rms fit drops to 0.5 and the relative rms fit decreases to 2.2%.

Figure 2 shows the STO-3G QM electrostatic potential and PD and M difference maps for water. For the M difference map, the rms fit is 11.3 and the relative rms fit is 40.4%. For the PD difference map, the rms fit drops to 0.5 and the relative rms fit drops to 1.6%. Figure 3 shows similar maps for the 6-31G** QM electrostatic potential; the rms fit of the M difference map is 5.5 and the relative rms fit is 16.6%. The rms fit of the PD difference map drops to 2.1 and the relative rms fit decreases to 6.1%. The STO-3G PD difference map for water shows that the model might be improved by the placement of small additional charges at approximately tetrahedral lone-pair sites. However, the 6-31G** difference map shows a shift of the excess positive charge to regions above and below the oxygen atom.

The ST2 empirical model potential of Stillinger and Rahman[9] for water assigns tetrahedrally located charge centers for the lone-pair electrons. Figure 4 shows maps of the ST2 electrostatic potential minus the STO-3G QM potential. The fit is not good, with an rms value of 10.4 and a relative rms fit of 39.1%. The fit to the 6-31G** potential is improved, showing an rms fit of 4.2 and a relative rms fit of 12.3%. However, the 6-31G** PD charges fit the potential still better. It should be noted that the PD charge model (and also the M charge model or any other atom-centered model) cannot reproduce the directional properties of hydrogen bonding in water clusters. Reproduction of this directionality was a principal objective in defining the ST2 model.

Figure 5 shows contour maps for ammonia. The M difference map shows an rms fit of 13.2, with a relative rms fit of 52.0%. In the PD difference map, these values are drastically reduced to 0.8 and 3.0%. The fit to the electrostatic potential by the PD charges is thus good enough so as not to require any consideration of the placement of a lone-pair charge site, to the indicated accuracy.

Figure 6 shows contour maps for methane; the M difference map has an rms fit of 0.8 and a relative rms fit of 49.3%. In the PD difference map these values are reduced to 0.2 and 7.8%.

Figure 7 shows maps for a triply bonded molecule, acetylene. The M difference map has an rms fit of 1.0 and a relative rms fit of 39.7. The PD difference map shows excellent agreement, with the rms fit dropping to 0.1 and the relative rms fit decreasing to 1.1%.

Figure 8 shows maps for the doubly bonded molecule, ethylene, using the STO-3G basis. The M difference map has an rms fit of of 1.4 and a relative rms fit of 34.6%. The PD difference map is not as good as that of acetylene; the rms fit decreases to 0.7 and the relative rms fit drops to 18.1%. Figure 9 shows similar maps using the 6-31G** basis. The PD difference map shows an improvement in its relative rms fit, with a value of 9.5%. The reason that the relative rms fit improves while the rms fit does not is because the 6-31G** electrostatic potential is generally larger in magnitude than the STO-3G potential (see Table I).

The 6-31G** PD difference map for ethylene

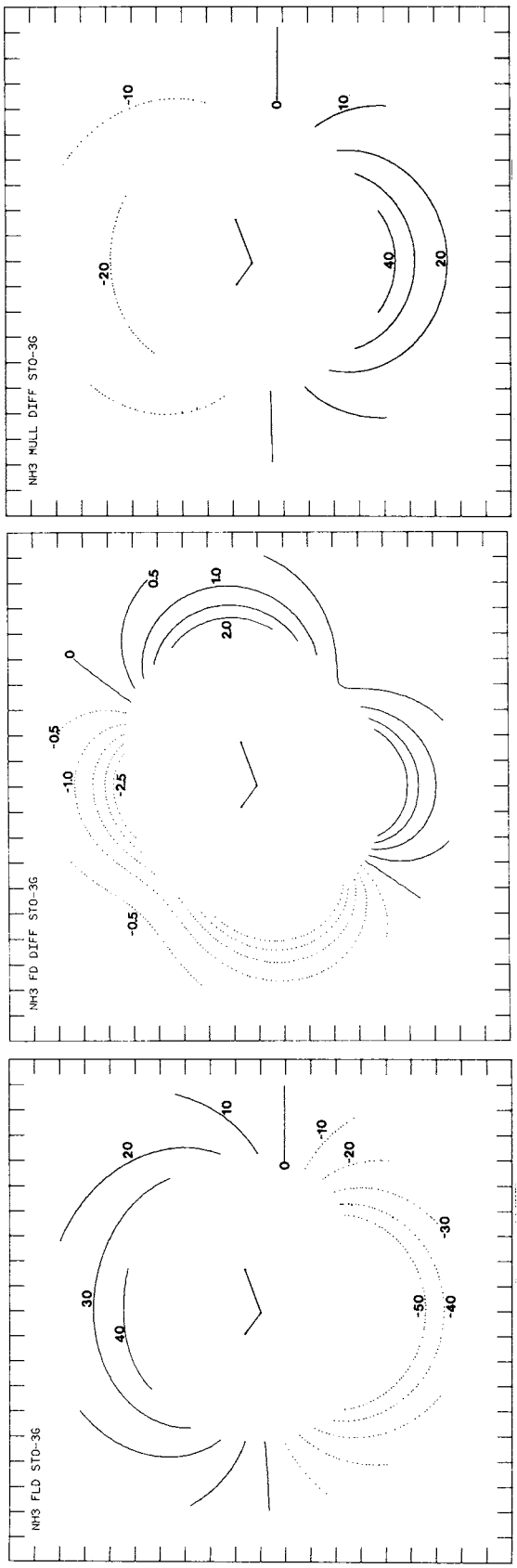

Figure 5. As in Fig. 1, for ammonia. The threefold axis of the molecule is vertical.

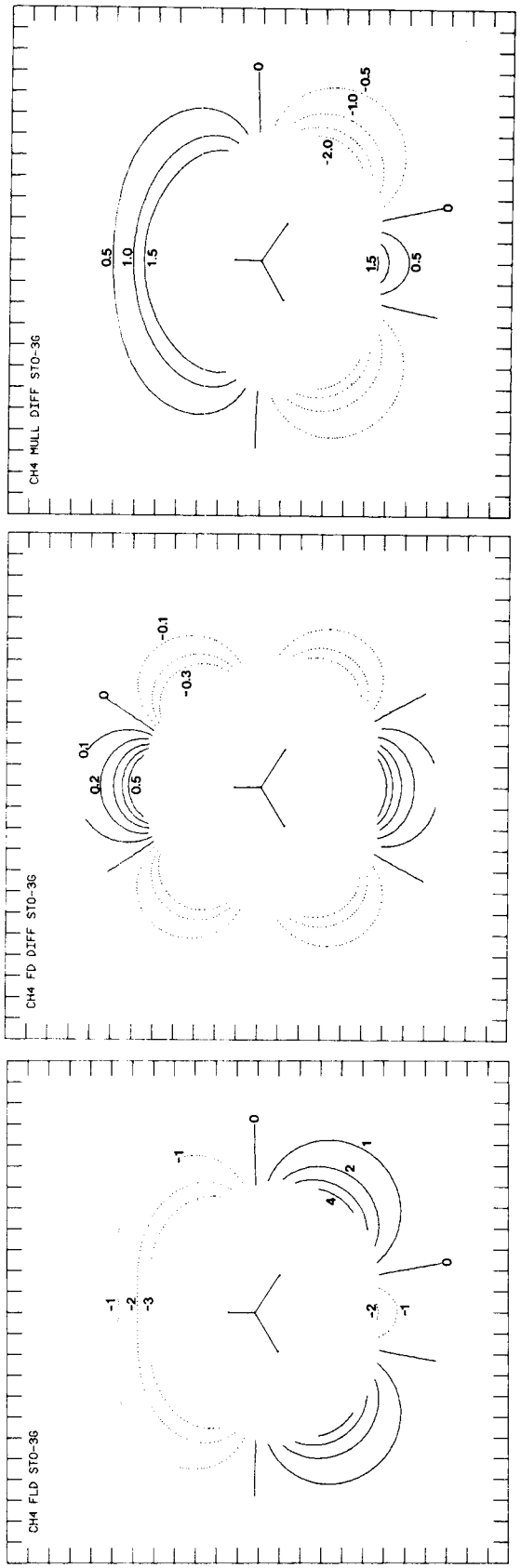

Figure 6. As in Fig. 1, for methane.

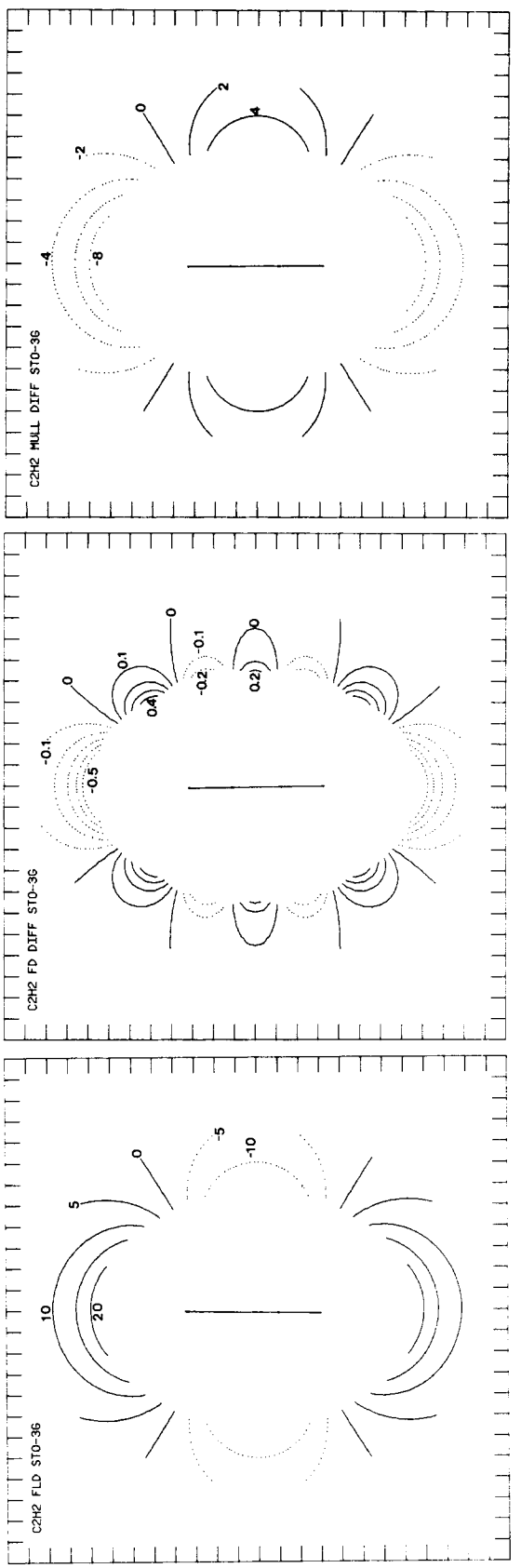

Figure 7. As in Fig. 1, for acetylene.

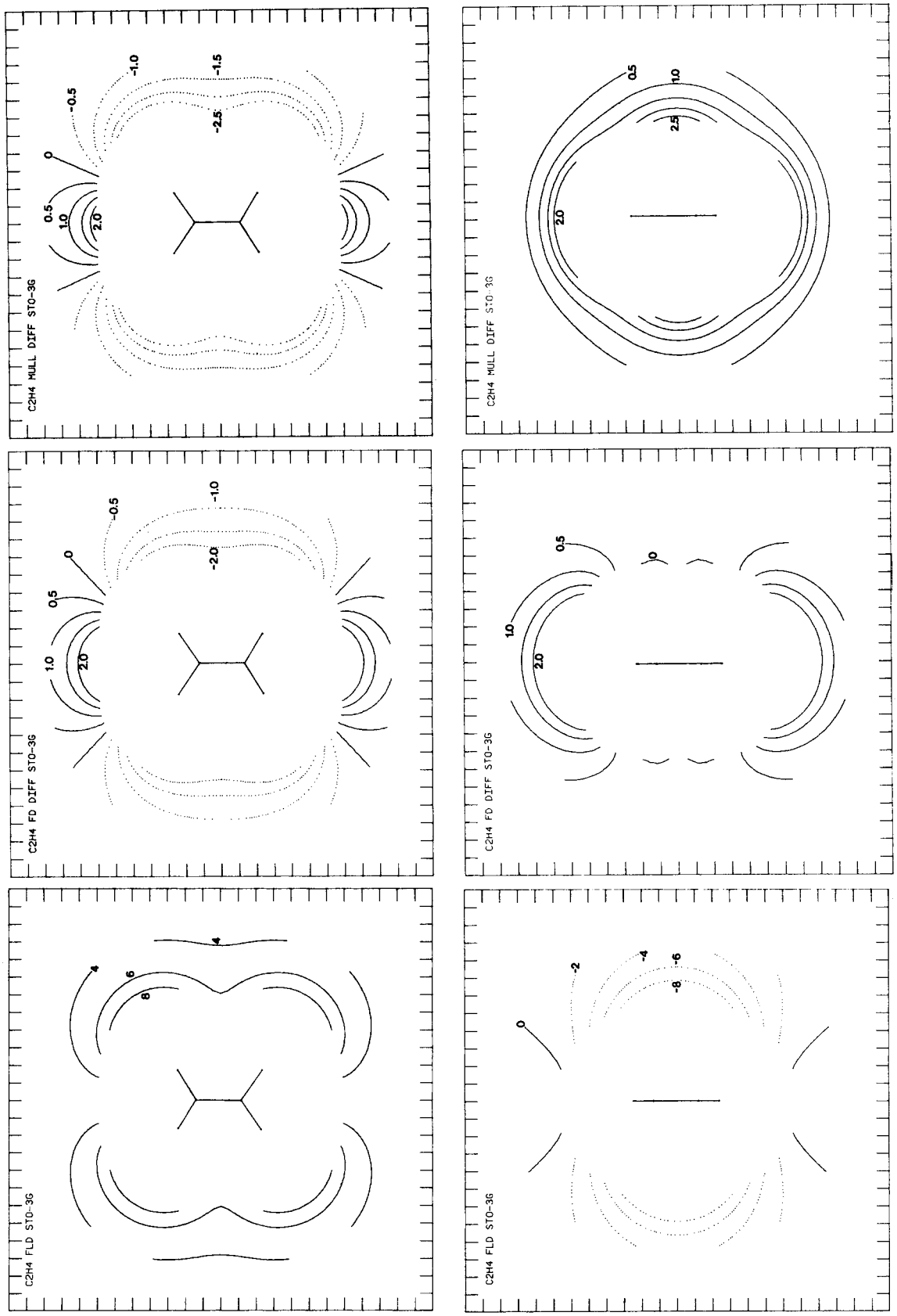

Figure 8. As in Fig. 2, for ethylene.

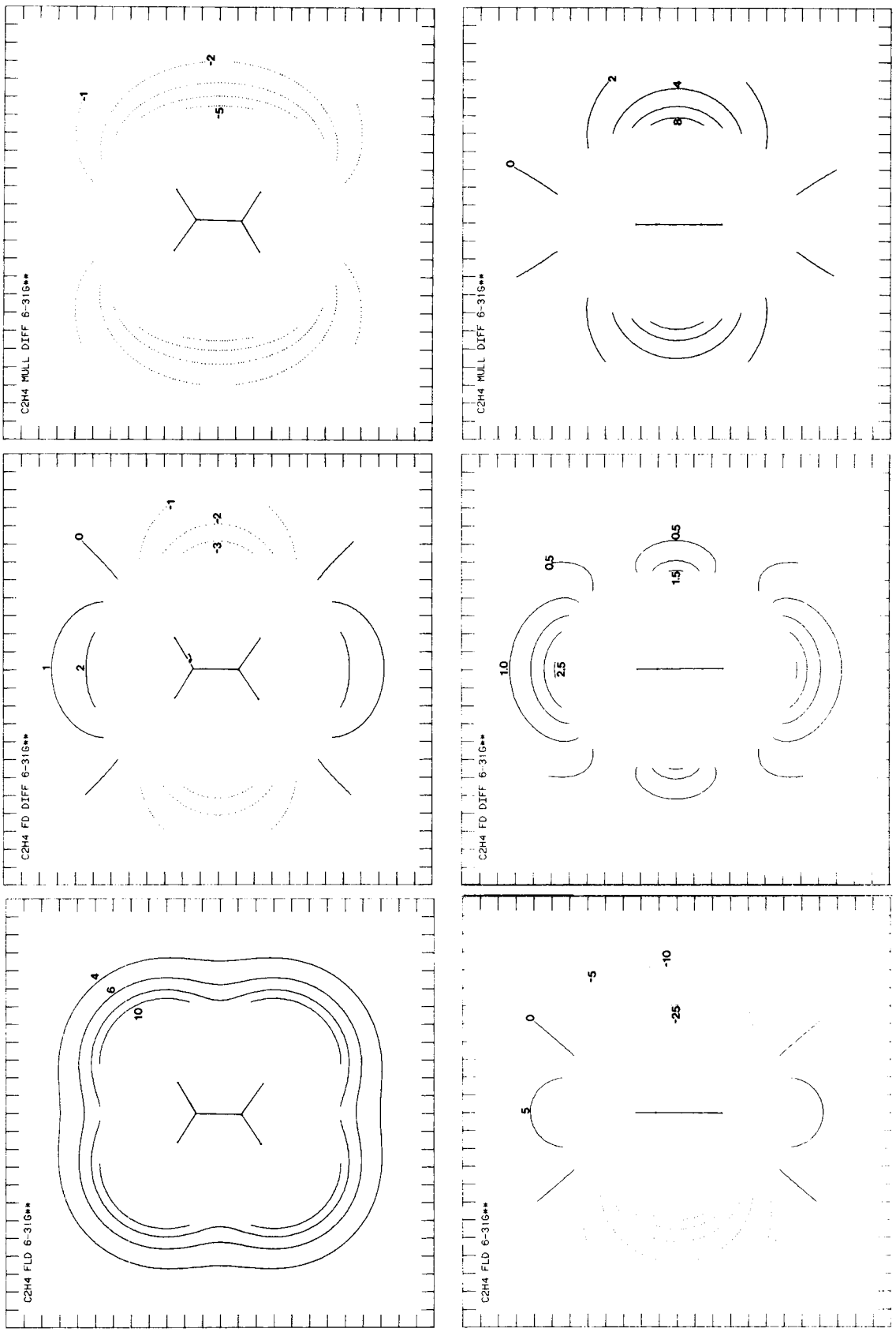

Figure 9. As in Fig. 3, for ethylene.

shows that the region of poorest fit is not perpendicular to the double bond; it is in the plane of the molecule. Wasiutynski, van der Avoird, and Berns[10] found that a net atomic charge model corresponding to a molecule with shrunken bond lengths gave better agreement to the calculated interaction energy between the molecules of ethylene dimer. We have also found that the PD model for the ethylene QM electrostatic potential improves if the molecule is shrunken by a similar amount. Although the relative rms fits for ethylene are rather high, the rms fits themselves are as good as for the other well-fitted molecules.

Figure 10 shows maps for carbon dioxide. The M difference map shows an rms fit of 5.9 and a relative rms fit of 42.6%, which is a poor fit. The PD difference map shows a much improved rms fit of 0.1 and a much better relative rms fit of 0.6%. We have found that cylindrically symmetric molecules usually give a very good PD difference map.

Figure 11 shows maps for formaldehyde. The M difference map shows a rms fit of 3.6 and a relative rms fit of 19.1%. The PD difference map shows an rms fit of 1.6 and a relative rms fit of 8.5%.

Figure 12 shows maps for formamide. Note that this molecule is not completely planar. The M difference map has an rms fit of 6.5 and a relative rms fit of 24.0%. The PD difference map shows a rms fit of 1.6 and a relative rms fit of 5.7%.

Momany[3] obtained PD charges and plotted electrostatic potentials for methanol, formic acid, and formamide. Our results for the PD charges are in good agreement for methanol and formic acid. However, there are some differences for formamide. Note that Momany used sampling points for the electrostatic potentials closer to the atoms than used in this work. The Momany charges for formamide show more positive oxygen and nitrogen atoms, while the carbon atom becomes more negative to preserve electrical neutrality.

Figure 13 shows maps for a molecule with bridging hydrogens, diborane. The M difference map has an rms fit of 3.8 and a relative rms fit of 73.7%. The PD difference map shows greatly reduced values of 0.4 and 7.2%. There does not seem to be any particular problem in fitting the QM potential near the bridging hydrogens.

Smit, Derissen, and Van Duijneveldt[11] obtained

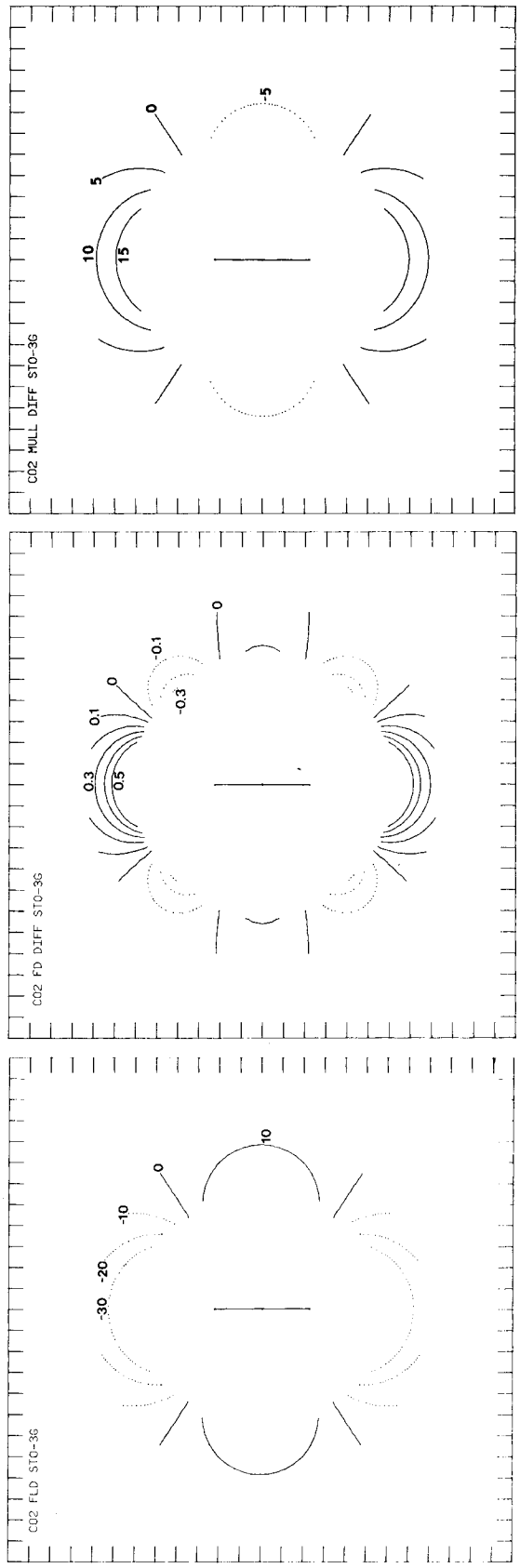

Figure 10. As in Fig. 1, for carbon dioxide.

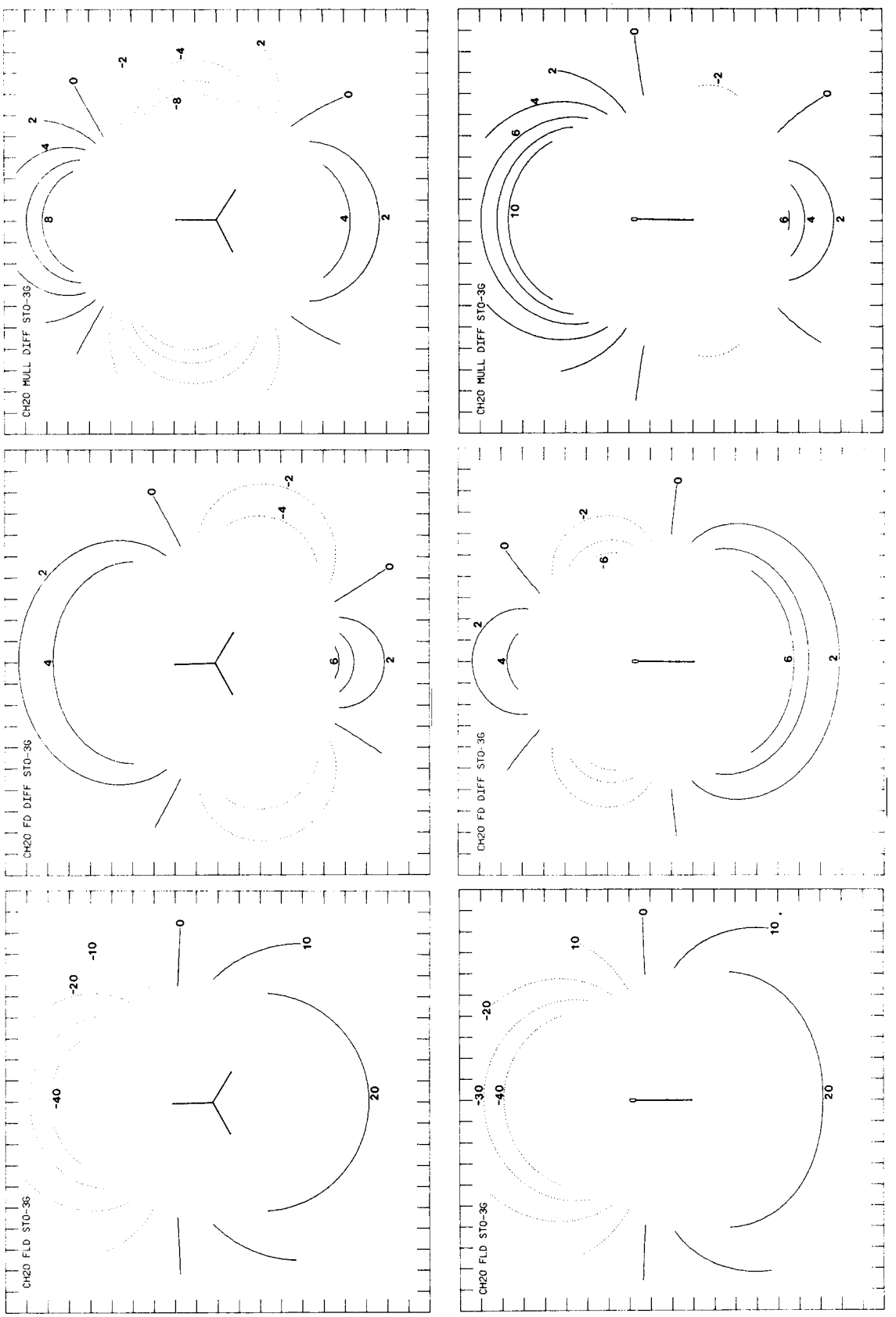

Figure 11. As in Fig. 2, for formaldehyde.

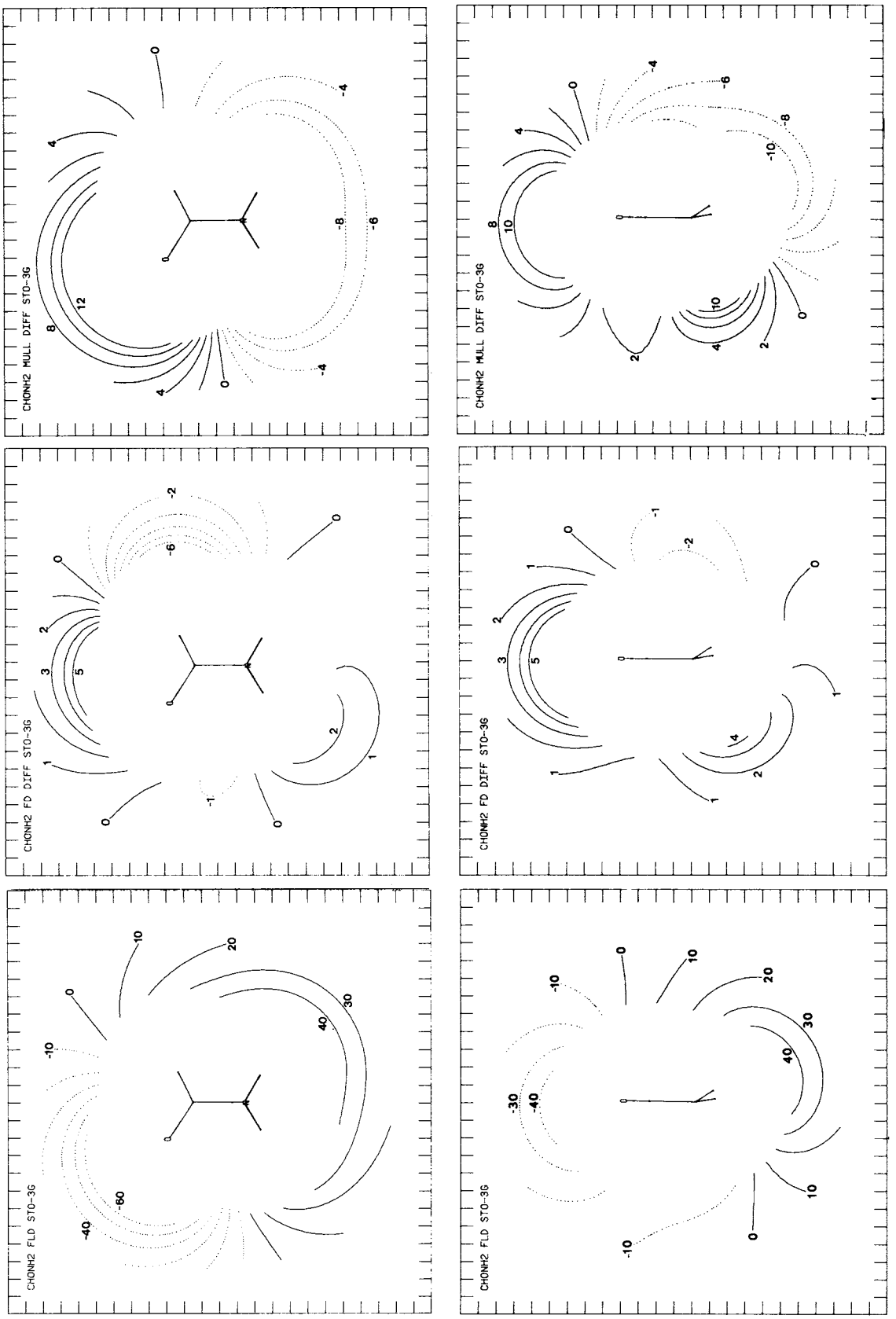

Figure 12. As in Fig. 2, for formamide.

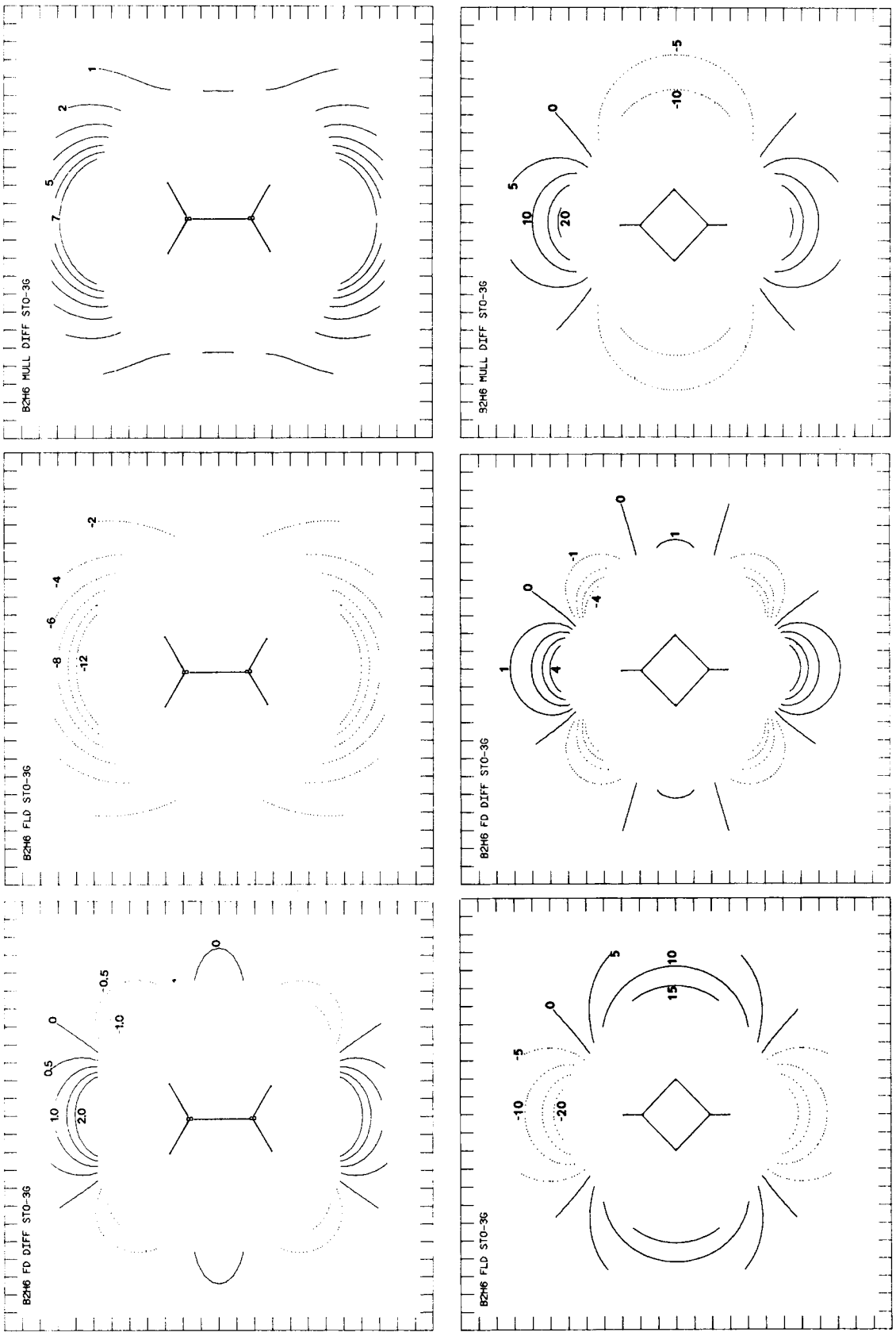

Figure 13. As in Fig. 2, for diborane.

Table VI. Comparison of net atomic charges in formamide.

Atom	Hagler, Huler, and Lifson[14]	PD 6-31G**
O	−0.383	−0.556
N	−0.829	−0.883
C	0.282	0.637
Carbon H	0.101	0.030
trans H	0.415	0.398
cis H	0.415	0.374

Table VII. Comparison of net atomic charges in formic acid.

Atom	Lifson, Hagler, and Dauber[15]	PD 6-31G**
Carbonyl O	−0.38	−0.628
Hydroxyl O	−0.38	−0.568
C	0.31	0.674
H	0.10	0.059
Hydroxyl H	0.35	0.462

PD charges for formaldehyde, methanol, and formic acid, using a split-valence basis wavefunction. Their results are similar to our PD charges using the 6-31G basis wavefunction.

COMPARISON WITH EXPERIMENTAL DATA

A comparison of the potential-derived charges can be made with charges optimized to give best fit to crystal structure data, which can be viewed as experimentally derived charges. It should be noted that net atomic charges derived from crystal structure data may show polarization effects from the surrounding molecules. Such polarization effects are not included in the simple PD charge model. Williams and Starr[12] obtained a charge of 0.153*e* on hydrogen attached to aromatic carbon, from a fit to observed crystal structures. If one makes the reasonable assumption that the charge for an aromatic hydrogen should be intermediate between hydrogens of single- and double-bonded carbons, then the average of the potential-derived charges for methane and ethylene should be close to 0.153*e*. In fact, this average for the 6-31G** basis is 0.158*e*, which represents excellent (although perhaps fortuitous) agreement. The STO-3G PD charges average to 0.102*e* which is less than this experimental value.

Yuen, Lister, and Nyburg[13] determined the net atomic charges on the carbonate anion by fitting the crystal structure of calcium carbonate (calcite). They obtained $q(C) = 0.95(26)$ and $q(O) = -0.98(9)$. The 6-31G** potential-derived charges (Table I) are $q(C) = 1.210(11)$ and $q(O) = -1.070(4)$, which are narrowly within the confidence limits of the crystal-structure-derived charges.

Hagler, Huler, and Lifson[14] derived a set of net atomic charges from the crystal structures of amides and from dipole moment data. They assumed as an approximation that the CO and NH_2 groups are electrically neutral, and that hydrogen bonded to carbon has the same charge as an alkane hydrogen. In the case of formamide, the HCO group was considered to be neutral, with an assumed charge of 0.101*e* on hydrogen. Table VI shows a comparison of the formamide net atomic charges obtained by Hagler, Huler, and Lifson with our PD net atomic charges. The agreement is seen to be fairly good for the amino group, but the PD charges show that the carbonyl group is considerably more polar than the Hagler, Huler, and Lifson model.

A similar approach was used by Lifson, Hagler, and Dauber[15] to obtain net atomic charges from the crystal structures and dipole moments of carboxylic acids. They found that the net charges on both the carbonyl oxygen and the hydroxyl oxygen could be taken equal and set to the same value found for the carbonyl oxygen in amides. This assumption left only one net atomic charge to be determined, that of the hydroxyl hydrogen; the carbonyl carbon charge was determined from the molecular neutrality condition. Table VII shows a comparison of the charges obtained by Lifson, Hagler, and Dauber with PD charges for formic acid. Again, the carbonyl group is shown to be considerably more polar when the PD charges are taken.

A further comparison can be made with the charges derived by Hirschfeld and Mirsky,[16] which were obtained by fitting the calculated molecular quadrupole moments of acetylene and carbon dioxide. For acetylene, the quadrupole charges found were $q(C) = -0.312$ and $q(H) = 0.312$; the 6-31G** PD charges are $q(C) = -0.297$ and $q(H)$

= 0.297, in good agreement. The STO-3G PD charges are smaller. For carbon dioxide, the quadrupole charges were $q(O) = -0.410$ and $q(C) = 0.820$, while the 6-31G** PD charges are $q(O) = -0.433$ and $q(C) = 0.867$, also in good agreement. In this case the STO-3G PD charges do not change much from the 6-31G** PD charges, and are also in good agreement with the cited values. There seems to be a pattern that STO-3G PD charges for hydrogen are too small, while the heavy-atom STO-3G PD charges agree better with the 6-31G** PD values.

In his work, Momany[3] also constrained the PD charges to reproduce the components of the observed dipole moment of the molecule. We believe that this procedure may not be beneficial for our purposes, since the calculated potential field describes the electrical nature of the molecule in detail in three dimensions more accurately than do the components of the dipole moment. We desire to represent this detailed electrical field as closely as possible with the PD charges. Since the net atomic charge model is not an exact one, our optimized net charges may not exactly fit the observed components of the dipole moment.

Another consideration is that the observed dipole moment may be zero by symmetry or may not have been experimentally measured. The lack of availability of an observed dipole moment would limit the generality of the dipole-fitting procedure. Similar remarks apply also to observed quadrupole moments and their fitting.

DISCUSSION

We were interested in the possibility of developing a prescription for obtaining realistic net atomic charges which would not require such excessive computer time as to prohibit application to large molecules. If we take the 6-31G** PD charges as reference values, we ask the question whether the STO-3G or 6-31G PD charges could be used as approximations. Table II shows that the STO-3G PD charges are smaller in magnitude than the reference values, but application of a scaling factor of 1.13 gives fairly good agreement with the reference values. The 6-31G PD charges are generally larger in magnitude than the reference values. A scaling factor of 0.82 applied to the 6-31G values yielded good agreement with the reference values (Table II). If these types of approximations are adequate, much computer time could be saved by calculating the STO-3G or 6-31G PD values and either using them directly or after scaling to approximate the 6-31G** PD values.

The authors thank Dr. J. L. Derissen and Dr. F. B. Van Duijneveldt for their critical reading of the manuscript. This work was supported by research grant GM-16260 from the National Institutes of Health.

References

1. E. Scrocco and J. Tomasi, *Adv. Quantum Chem.*, **11,** 115 (1978).
2. R. S. Mulliken, *J. Chem. Phys.*, **23,** 1833 (1955).
3. F. A. Momany, *J. Phys. Chem.*, **82,** 592 (1978).
4. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51,** 2657 (1969).
5. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **28,** 213 (1973).
6. P. A. Marsh and D. E. Williams, *Quantum Chemistry Program Exchange*, University of Indiana, Bloomington, IN, 1981 (to be published).
7. C. W. Kern and M. Karplus, *J. Chem. Phys.*, **40,** 1374 (1964).
8. L. C. Snyder and H. Basch, *Molecular Wave Functions and Properties: Tabulated from SCF Calculations in a Gaussian Basis Set*, Wiley, New York, 1972.
9. F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **60,** 1545 (1974).
10. T. Wasiutynski, A. van der Avoird, and R. M. Berns, *J. Chem. Phys.*, **69,** 5288 (1978).
11. P. H. Smit, J. L. Derissen, and F. B. Van Duijneveldt, *Mol. Phys.*, **37,** 521 (1979).
12. D. E. Williams and T. L. Starr, *Comput. Chem.*, **1,** 173 (1977).
13. S. P. Yuen, M. W. Lister, and S. C. Nyburg, *J. Chem. Phys.*, **68,** 1936 (1978).
14. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96,** 5319 (1974).
15. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101,** 5111 (1979).
16. F. L. Hirschfeld and K. Mirsky, *Acta Crystallogr. Sect. A*, **35,** 366 (1979).