Theor Chim Acta (1986) 69: 71–81

Theoretica
Chimica Acta

The electron density of the water molecule

G. G. Hall and C. M. Smith

Division of Molecular Engineering, Kyoto University, Kyoto 606, Japan

(Received June 27, revised September 11/Accepted September 12, 1985)

The electron density of the water molecule, as calculated by a standard program, is approximated by linear combinations of spherical Gaussians. The accuracy of the result is studied as a function of the numbers and positions of the Gaussians. Since this shows where the charge is located in the molecule it has immediate physical significance. The building-up of the density can be followed in more and more detail. From these expansions, point charge models of water are readily deduced. These are compared with models of similar kinds used by other authors. Some of the calculations have been repeated with a wavefunction of higher accuracy to investigate the stability of the results. Results show that the more accurate density requires more Gaussians to represent its greater complexity.

Key words: Electron density — Gaussian models — Point charge models — Water molecule

1. Introduction

Now that advances in computing have made it possible to calculate many wavefunctions for molecules there is an increasing need for methods of summarizing and comparing the results so that the insight which they contain into the structure of the molecule can be fully exposed. This paper is concerned with two examples of this. It demonstrates a method of analysing the electron density in such a way as to display clearly its important features. It also derives a point charge model which has certain advantages in the calculation of particular molecular properties such as the molecular electrostatic potential and its derived electric field.

The Hohenberg-Kohn [1] theorem has produced a major change in the interpretation of theoretical chemistry. In the past it was thought essential to analyse the electron correlation function in order to understand chemical bonding etc. Now it is clear that a full understanding of the electron density is sufficient to explain all the properties which the molecule has in a particular state. Comparisons of electron densities using difference maps have clarified many features of bonds in small molecules but are very limited in application to larger systems. Numerical comparisons generalize more readily. One such entity, and perhaps the only one in regular use, is the Mulliken population [2]. A recent review [3] discusses the defects of this definition and of various attempts to improve it. This paper uses a different kind of analysis of the electron density which is more flexible and more accurate. It fits the electron density by a linear combination of Gaussians centered on different points, using a fitting functional which is the square of the error in the electric field [4]. Such an approximation can be very compact compared with the original density and so easier to interpret. Details of the method of calculation are given in the preceding paper [5].

Compact representations of the electron density can be used for many purposes. In particular, the molecular electrostatic potential (MEP) at any point is more easily calculated. This quantity dominates the long-range intermolecular forces and so exerts a major influence on chemical reactivity and molecular beam scattering. The MEP can be calculated very rapidly using point charge models to represent the electrons. A very simple method exists of deriving such models from the approximate Gaussian models [6].

To demonstrate the utility of these methods in detail a calculation on the water molecule is reported. The major terms in the electron density are identified and located. By repeating the calculation with a more accurate wavefunction the sensitivity of the results to the complexity of the wavefunction can be judged. Point charge models of different levels of accuracy are derived and compared with those used by other authors. Many authors have started with some concept of the density and used it as a model the parameters of which were determined by fitting selected empirical data. The adequacy of their ideas as well as the accuracy of their models can now be tested.

2. The atomic models

The simplest concept of a molecular density is as a sum of spherically-symmetrical densities centered on each atom of the molecule. This model is adequate enough for some purposes. The idea behind the definition of atomic populations is also that of spherical densities and this becomes explicit when populations are used to define point charge models [7]. On the other hand the concept ignores significant aspects of bonding so its accuracy is limited. In this section these limits will be explored.

The simplest model involving only the three nuclear centers would have three Gaussians. This is so bad a representation of the nuclear cusp that it was not used. With two Gaussians on the O nucleus the density begins to concentrate in the inner region and the result is poor but sensible. The use of more Gaussians improves the representation of the valence electron densities and also gives sharper densities at the nuclei (i.e. uses larger exponents). The details of the exponents and charges for each Gaussian in two of these models, the simplest and the most elaborate, is given in Table 1. It is convenient to label these Gaussian models by (r, s), where r is the number of centres and s the total number of Gaussians. The charges for all Gaussian models are listed as electron charges whereas for point charge models proton charges are used. All distances quoted are in Å.

Table 1 shows how the extra flexibility of the (3, 9) model is used to improve the shape of the density in each atom and how the electronic charge is distributed among the Gaussians on each atom. The total electron charge on each atom does not change very much. Figure 1 shows the error integral (i.e. U, in the notation of [5]) for each $(3, n)$ model as a function of n, the total number of Gaussians. The models in this section are in the sequence with label 3. It can be seen from this that there is a significant improvement in the models on going from 6 to 7 Gaussians and this means giving the O a third Gaussian so that two represent the K shell and one the valence electrons. Of course, for the density these categories are not so precise as in an orbital picture and the charges are not close to integers but the interpretation seems to remain good. Figure 1 also shows that the improvement from 8 to 9 Gaussians is very small so that this is close to the limit of accuracy obtainable using three centres and spherical densities.

To derive the point charge models from these Gaussian models each Gaussian is shrunk to a delta function [6]. This process preserves all the spherical moments (for a more general proof of this see [8]) so all the angular dependence is retained. The two point charge models from the Gaussian models above are shown in Table 2. A model from previous calculations [9] using the Snyder-Basch [10] wavefunction is given to show the effect of using a different starting wavefunction. Two models derived by fitting the electron density [11, 12] show the sensitivity of these results to different versions of this criterion. Clementi and his colleagues have used the long-range forces between dimers to fix their atomic charges and the position of the O charge, which moves on the axis towards the H atoms.

Table 1. The exponents and charges of the Gaussians in two models of water using the nuclei only as centers. The charge is in electrons

Label	Atom	Exponent	Charge	Total charge on atom
(3, 4)	O	1.05995	6.80785	
		52.84929	1.68505	8.49290
	H	1.01999	0.75355	0.75355
(3, 9)	O	0.68212	2.99043	
		1.36258	3.86689	
		31.97357	1.21855	
		104.68234	0.49838	
		508.18792	0.04971	8.62396
	H	0.33115	0.31881	
		2.67226	0.36921	0.68802

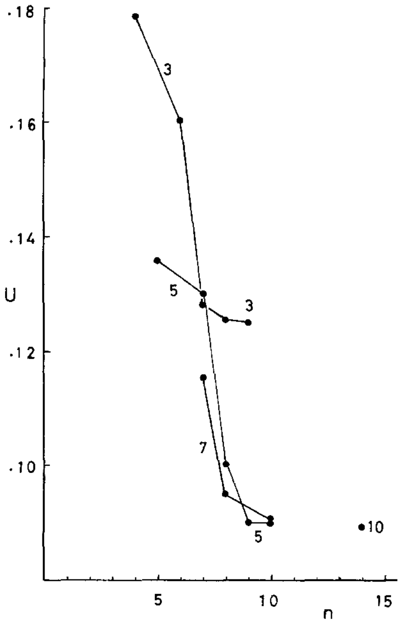

Fig. 1. The error integral, *U*, (in hartrees) for Gaussian models of water as a function of *n*, the total number of Gaussians. The label denotes the number of centres. The original electron density is from a STO-6G wavefunction

They distinguish in their results between calculations without correlation [13] and with correlation [14, 15]. Also in this table are the similar models, given by Brobjer and Murrell [16] and by Kollman [17], which use the experimental dipole moment. (It should be noted that this is the simplest of the Brobjer and Murrell models. Their more adequate models do not compare so directly with those here.) The model used by Ray et al. [18] is deduced by fitting the MEP. It can be seen that all are alike in making the O negative and by about the same amount but the numbers do not agree because of their very different sources.

It should be noted that these models do not refer to exactly the same geometry though this does not have a large effect on the results.

Table 2. Point charge models of water using nuclear charges only. These charges are in protons

Net O charge	Net H charge	Source
−0.492	0.246	(3, 4)
−0.62396	0.31198	(3, 9)
−1.167	0.584	[9]
−0.593	0.296	[11]
−0.614	0.307	[12]
−1.34072	0.67036	[13]
−1.50348	0.75174	[15]
−0.6574	0.3287	[16]
−0.658	0.329	[17]
−0.56	0.28	[18]

3. The lone pairs

There are several possible positions for an extra charge outside the spherical atoms of the previous section. By trying alternatives, it was shown that the lone pair region, as might be expected, is the most important. Models with some Gaussians centered in the lone pairs, 5-center models, should be more accurate and more useful than the simpler 3-center ones above.

Table 3 shows the details of three $(5, n)$ models, each having one Gaussian on each lone pair (denoted by l) and two on each of the H nuclei.

This table shows several dramatic effects. Whereas in the (5, 8) model each lone pair charge is more than 3 electrons, in the more accurate models it reduces to about 1/2 electron. To compensate, one of the O functions which is not represented in the (5, 8) model, begins to collect over 5 electrons. This clearly illustrates the danger of drawing conclusions too soon! The asymmetry due to the lone pairs is important but it is exaggerated in the first function because there is not sufficient flexibility on the O to represent the valence electrons. The later functions have three or four O functions so the lone pair functions do represent the genuine lack of spherical symmetry and their effect is small but significant.

This kind of major change in the location of charge, as the number of Gaussians is increased, is often associated with the presence of a diffuse Gaussian in the model. These functions, having exponents smaller than 1, are spread over the molecule and should not be regarded as belonging exclusively to one part of it. They compete with the more localized Gaussians in describing the total charge

Table 3. Exponents and charges for models on nuclei and lone pairs. The lone pair-O distance is $|Ol|$ and is in Å

Label	Site	$\|Ol\|$	Exponent	Charge	Total charge on site
(5, 8)	O		34.37057	1.40037	
			163.98346	0.37804	1.77841
	l	0.174	1.22218	3.15136	3.15136
	H		0.47361	0.47624	
			2.12756	0.48319	0.95943
(5, 9)	O		0.89233	5.37491	
			38.53273	1.41504	
			189.04908	0.29746	7.08741
	l	0.236	3.20707	0.66031	0.66031
	H		0.46404	0.35667	
			2.12756	0.43931	0.79598
(5, 10)	O		0.93507	5.73263	
			28.05798	0.72038	
			55.63883	0.84340	
			256.32041	0.18178	7.47819
	l	0.257	4.08644	0.45766	0.45766
	H		0.39547	0.35471	
			2.12756	0.44853	0.80324

in each region and so provide a coupling between the localized charges in the fitting equations. This can lead to the major changes in the charges which is found when a new function is added. In our previous paper [9] it has been shown that by partitioning the Gaussians, and especially the diffuse Gaussians, after the manner proposed by Bader [19] the worst effects of these changes can be removed. The atomic charges and dipoles are then defined by integration over the space assigned to each atom. This produces a partitioning of each Gaussian charge. The atomic charges and dipoles so defined are remarkably stable, even when the Gaussian charges suffer major changes during the improvement of the fit, and justify their use as general definitions of these quantities. Thus, for example, the charge on O for the third fitting listed in Table 2 is reduced from -1.167 to -0.650 which is much closer to the values in the remaining models. The charges listed here have not been partitioned and are less precise in their interpretation of the location of the charge density components, since all the charge on a diffuse function is located at its center. On the other hand, they have the major advantage of relating to spherical densities and leading directly to a point charge model involving the location of charges on or off the nuclei. The Bader atomic density loses spherical symmetry because of its spatial boundary and so many atomic moments are required to describe it. Its charges and moments are more easily related to the "segmented multipole moment" [20] models which have charges and multipoles on the nuclei.

In these models the most diffuse Gaussian is on each H atom. Its exponent changes as the number of Gaussians on O changes and so does its charge. In contrast, the more localized Gaussian on H optimizes with exactly the same exponent and very similar charge.

The small charges in these models are also of interest. They indicate the size of the smallest amount of charge which has been successfully allowed to find its correct location during the optimization. In these models this size has reduced to 0.18 electron. This gives a crude, but ready measure of the fitting error.

The errors of these $(5, n)$ models are shown in Fig. 1 with the label 5. It can be seen that the error has been significantly reduced by the extra flexibility due to these lone pair Gaussians. This reduction is a measure of the asymmetry in electron density around the O. The large drop in the error when the O acquires 3 Gaussians is also seen in both $(3, n)$ and $(5, n)$ sequences.

Point charge models of the water lone pairs have been widely used in various chemical contexts especially when some representation of the hydrogen bond was required. Some are listed in Table 4 along with those derived from the models above.

With the exception of the $(5, 8)$ model, which is the least accurate one, these models show broad agreement. The H charge is seen to be small and positive. The lone pair charge is fairly small and the net remaining O charge is so small that even its sign is not agreed. The largest difference is in the O-l distance but, since the other authors had no direct means of determining this, some differences

Table 4. Point charge models of water including lone pairs

$\lvert Ol \rvert$	Net O charge	*l* charge	Net H charge	Source
0.174	6.2216	−3.1514	0.0406	(5, 8)
0.236	0.9126	−0.6603	0.204	(5, 9)
0.257	0.5218	−0.4577	0.1968	(5, 10)
0.8	0	−0.2357	0.2357	[21]
1.4	−0.300	−0.078	0.228	[17]
0.7	−0.144	−0.156	0.228	[17]

must be expected. It is more significant that this distance multiplied by the lone pair charge i.e. the lone pair moment, does not vary by much.

4. The OH bond

The Gaussian models above effectively ignore any lack of spherical symmetry around the H nucleus. A diffuse function on the O will produce a small asymmetry around the H but such functions are influenced by the whole molecule rather than any one part of it so that this may not be significant. If the electron density is to express the effect of bonding then some change of shape around H must be expected. In this section some models of the OH region will be described. In these models the accuracy is increasing but so is the complexity of the model.

Table 5 gives the details of three models which include these contributions in the OH bond. For convenience a Gaussian along the bond will be denoted as *c*.

Table 5. Gaussian models including lone pair and OH bond functions

Label	Site	$\lvert Ol \rvert$ or $\lvert Oc \rvert$	Exponent	Charge	Total charge
(7, 7)	O		58.27035	1.55603	
	l	0.193	3.19711	0.93770	
	c	0.031	0.78683	2.59564	
	H		1.37472	0.68865	
(7, 10)	O		0.93507	5.40179	
			28.05798	0.72842	
			55.63884	0.83690	
			256.32040	0.18262	7.14973
	l	0.260	4.08644	0.49417	
	c	0.612	0.39547	0.47557	
	H		2.12756	0.45540	
(10, 14)	O	0.102	0.35465	−0.01693	
			31.54198	1.11408	
			83.89791	0.54027	
			345.83829	0.09948	1.73690
	l	0.262	3.36363	0.56271	
		−0.023	0.92651	2.73678	
	c	0.821	0.30021	0.20213	
	H		0.59650	0.19726	
			2.12756	0.43268	0.63094

The O-c distance is then $|\mathrm{O}c|$. There are several interesting features in this table. The third model is an example where the charge changes sign on one Gaussian. The significance of this is discussed in the preceding paper [5]. In this model one of the O Gaussians has been allowed to float off the nucleus and it has moved slightly into the lone pair region but is still in the molecular plane. It has a small positive charge. This model also has a second, more diffuse function to represent the lone pairs but it has floated through the nucleus to the H side of the O nucleus. Such a detailed model of the O valence region makes it more likely that a positive charge will arise.

The (7, 7) model has a small charge centered on O. It is essentially the K shell. The OH charge is quite large since it includes the valence shell on that side of O and balances the lone pair charge on the other side. In the (7, 10) model, however, the O charge swings up again due to the extra O function which can represent the valence shell more adequately. The (10, 14) model provides another option, a second Gaussian for each lone pair, and this suits the valence shell even better. Although some of the changes in charge are large, they are more apparent than real since the charge is sometimes moving only a short distance and some functions are quite diffuse.

This table includes an example of the process, described in [5], of developing a model in small stages. The Gaussians used in the model (7, 10) have the same exponents as those in model (5, 10). The only difference is that a H function has been allowed to float into the bond. The redistribution of charge is not large but it is significant. This model is probably at a local minimum and so further improvement in fitting can be obtained by a reoptimization from a new starting point.

The error in these models is plotted in Fig. 1 as the sequence labelled 7 except for the third function which is the point 10. It can be seen that increasing the number of centers in these models is not always beneficial since, for the same number of Gaussians, some of the 5 sequence models are more accurate. The broad conclusion from these models is that improving the lone pair region and putting at least two Gaussians on H is more important than having a Gaussian in the OH bond but that it does then give an improvement. The (10, 14) model above is the most accurate one calculated in this paper.

Table 6. Point charge models using OH functions

$\lvert \mathrm{O}l \rvert$	l charge	Net O charge	Net H charge	c charge	$\lvert \mathrm{O}c \rvert$	Source
0.193	−0.9377	6.4440	0.3113	−2.5956	0.031	(7, 7)
0.260	−0.4942	0.8503	0.5446	−0.4756	0.612	(7, 10)
0.262	−0.5627	6.2631	0.3691	−0.2021	0.821	(10, 14)
0.023	−2.7368					
0.297	−2	6	1	−2	0.57	[22]

The point charge models derived from these Gaussian models are shown in Table 6. This also includes a model due to Shipman et al. [22] in which the charges are taken to be integers and the two position parameters are determined using experimental properties of ice. It is interesting to note that, for such a type of model, the charges calculated here are not greatly different from theirs.

5. Comparisons with other wavefunctions

In our earlier paper [9] water was included as one example. The original wavefunction was the Snyder–Basch function [10] which is a relatively simple function. Although only Gaussians on the nuclei were used, the approximate density converged to a very low error after a modest number of terms, 6 or 7 Gaussians being sufficient to reduce the error to 0.01. Since the total self-energy is about 47 hartree, this represents a high accuracy. In this paper the wavefunction is a STO-6G one which has many more primitive Gaussians in the orbitals and a much more complicated density function. The result is that the fitting has a slower convergence and, with a limited number of terms, gives a larger limit. The best fit here has an error of 0.089.

To investigate further the sensitivity of the results to the form of wavefunction, calculations were repeated using a 4-31G* basis, which includes polarization functions. Table 7 shows details of one of these models having 9 centers.

This model is easily related to those above. It has two Gaussians for its *K* shell and represents the valence electrons through the lone pair functions. Since it is a (5, 9) model, it directly compares with the previous (5, 9) model in Table 3. Perhaps the most significant difference between them is the change in the description of the lone pair. Although 10 Gaussians are used the accuracy of this fit is rather low. The errors of all the models calculated from this wavefunction are shown in Fig. 2. Even the smallest error is twice as large as for the STO-6G wavefunction used earlier and the sequences show little sign of convergence. The effect of the polarization functions is to introduce more non-spherical character into the atomic densities. Since the models used here have a limited number of spherical Gaussians this departure from the local spherical symmetry cannot be fully represented and so the error is larger. Thus the pattern discerned before, of a more complex wavefunction being more difficult to fit, is repeated here.

Table 7. A Guassian model based on a polarization wavefunction

Label	Site	$\lvert Ol \rvert$	Exponent	Charge	Total charge
(5, 9)	O		0.74286	5.16919	
			36.21397	1.40124	
			181.44297	0.32630	6.89673
	l	0.133	2.70911	0.86295	0.86295
	H		0.24633	0.22688	
			1.88290	0.46181	0.68869

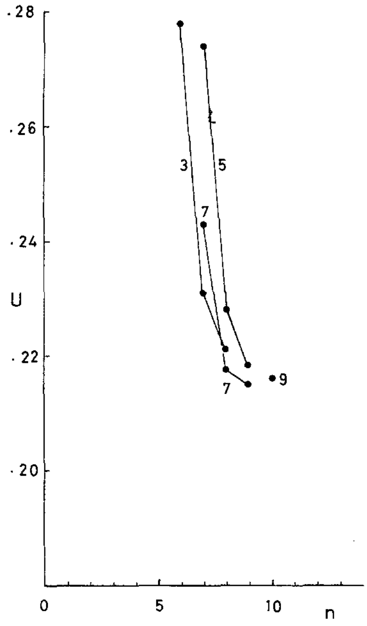

Fig. 2. The error integral for Gaussian models similar to those in Fig. 1 but based on a 4-31G* wavefunction

6. The molecular electrostatic potential

One of the principal uses of point charge models is in making rapid calculations of the MEP. This is especially needed in molecular dynamics simulations where the speed of calculation is a limiting factor. Many of the point charge models quoted above were developed for this purpose.

A major feature of the MEP is the presence of a minimum in the outer parts of some molecules. This occurs in lone pair regions and also in some double bonds and bent bonds. Clearly, it is important in estimating the intermolecular forces with a proton or when an electropositive region in another molecule is concerned. Representing such a feature is not easy. Earnshaw's theorem [23, 8] says that a potential cannot have a minimum except at a point occupied by a negative charge. Thus models which use the nuclei alone, even though they have several multipoles there, will fail to represent this feature. On the other hand a point charge model which has a charge in the lone pair, can do so. This is the principal reason why, in this paper, models are used which include points off the nuclei rather than having a variety of multipoles on them.

There is another advantage in moving some functions off-center. When molecules come close the asymptotic form for the forces is no longer adequate to express their interaction. Penetration effects begin to play an important role. These are significant when the molecules begin to be involved in bonding, as in some applications to chemical reactions. A set of point charges can be used to represent these penetration effects [4]. In the simple point charge models of this paper penetration is not well represented but, because the error is taken over all space, it is included in the Gaussian models and so does have some effect on the point charge models. These models should then remain valid at shorter intermolecular distances than models that use only the nuclear charges and multipoles. At small intermolecular distances the MEP alone is not sufficient to represent the interaction so the models should not be applied then.

7. Conclusion

This paper has demonstrated the analysis of a wavefunction through its electron density. It has located the concentrations of charge and shown alternative ways of representing them. It has also deduced point charge models from *ab initio* wavefunctions in a rigorous way instead of using arguments the accuracy of which can be debated. In particular it offers point charge models which can describe the lone pairs of molecules much more accurately.

Acknowledgement. The authors wish to thank the Japanese Ministry of Education for a Scholarship (to CMS).

References

1. Hohenberg, P., Kohn, W.: Phys. Rev. **136**, B864 (1964)
2. Mulliken, R. S.: J. Chem. Phys. **3**, 573 (1935)
3. Hall, G. G.: Adv. Atomic Mol. Phys. **20**, 41 (1985)
4. Hall, G. G.: Theoret. Chim. Acta (Berl.) **63**, 357 (1983)
5. Smith, C. M., Hall, G. G.: Theoret. Chim. Acta **69**, 63 (1986)
6. Hall, G. G.: Chem. Phys. Lett. **20**, 501 (1973)
7. Mulliken, R. S.: J. Chem. Phys. **23**, 1833 (1955)
8. Hall, G. G.: Int. Rev. Phys. Chem., in press
9. Hall, G. G., Smith, C. M.: Int. J. Quant. Chem. **25**, 881 (1984)
10. Snyder, L. C., Basch, H.: Molecular wave functions and properties. New York: Wiley 1972
11. Yanez, M., Stewart, R. F., Pople, J. A.: Acta Crystall. **A34**, 641 (1978)
12. Escudero, F., Yanez, M.: Mol. Phys. **45**, 617 (1982)
13. Popkie, H., Kistenmacher, H., Clementi, E.: J. Chem. Phys. **59**, 1325 (1973)
14. Matsuoka, O., Clementi, E., Yoshimine, M.: J. Chem. Phys. **64**, 1351 (1976)
15. Carravetta, V., Clementi, E.: J. Chem. Phys. **81**, 2646 (1984)
16. Brobjer, J. T., Murrell, J. N.: J. Chem. Soc. Faraday Trans. 2 **78**, 1853 (1982)
17. Kollman, P. A.: J. Am. Chem. Soc. **100**, 2974 (1978)
18. Ray, N. K., Bolis, G., Shibata, M., Rein, R.: Int. J. Quant. Chem. **QB11**, 257 (1984)
19. Bader, R. W. F., Tal, Y., Anderson, S. G., Deng, T. T. N.: Isr. J. Chem. **19**, 8 (1980)
20. Sokalski, W. A., Poirier, R.: Chem. Phys. Lett. **98**, 86 (1983)
21. Stillinger, F. H., Rahman, A.: J. Chem. Phys. **60**, 1545 (1974)
22. Shipman, L. L., Burgess, A. W., Scheraga, H. A.: Proc. Natl. Acad. Sci. **72**, 543 (1975)
23. Stratton, J. A.: Electromagnetic theory. New York: McGraw-Hill 1941