A new orbital-based model for the analysis of experimental molecular charge densities: an application to (Z)-N-methyl-C-phenylnitrone

David E. Hibbs,[*a] **Siân T. Howard,**[*a] **Jeremy P. Huke**[b] **and Mark P. Waller**[a]

[a] *Faculty of Pharmacy, University of Sydney, NSW, 2006, Australia.*
  E-mail: davidh@pharm.usyd.edu.au, sian@pharm.usyd.edu.au
[b] *Department of Mathematics, UMIST, Manchester, UK M60 1QD*

An alternative to the usual atom-centred multipole expansion is presented for the analysis of high resolution, low-temperature X-ray scattering data. The molecular electron density is determined in a fixed basis of molecular orbitals with variable orbital occupation numbers, *i.e.* the same form which is used to represent the density in *ab initio* electron-correlated calculations. The advantages of such an approach include linear scaling (in the sense that the number of parameters to be determined by fitting varies linearly with system size) and ease of property calculation. The method is applied to experimental high-resolution structure factors for a phenylnitrone, and compared to the results of a multipole model of the same data. Finally, the model is critically compared with several related, published orbital-based models.

Introduction

The experimental determination of molecular charge distributions from high-resolution, low-temperature elastic X-ray scattering experiments is now a mature field of science, as described in a number of recent reviews.[1,2] Conventional analysis of observed X-ray structure factors $\{F_o(s)\}$, which are the amplitudes of diffracted radiation at N "lattice points" $s = ha^* + kb^* + lc^*$ in reciprocal space which obey the Bragg conditions for a given crystal, depends on the comparison with a set of model structure factors $\{F_c(s)\}$ (eqn. (1)) where $f_k(s)$ and $T_k(s)$ are respectively the "atomic" scattering factors and temperature factors of the k-th atom.

$$F_c(s) = \sum_{k=1}^{N_{atoms}} f_k(s) T_k(s) \exp(2\pi i\, s \cdot r_k) \qquad (1)$$

The difference between a standard crystallographic application of eqn. (1) and the charge density analysis using multipole models lies only in the form of the scattering factors. In the former, so-called independent atom model (IAM) these are fixed, spherically-symmetric functions (the Fourier transforms of gas-phase (usually ground state) free atom/ion densities); in the latter, these are a set of aspherical, flexible functions (the Fourier transforms of real space charge density multipoles, see ref. 1–2 for details). Although this extra flexibility introduces more parameters into the fitting process, the multipole model is still essentially "linear-scaling" in the sense that the number of parameters to be determined scales linearly with the molecule size. Moreover, the observations/parameters ratio obtained for typical multipole analysis of most high-resolution data sets is high enough (say 10 or better) to guarantee a unique and rapidly-convergent fit. Principally for these reasons, and despite being fundamentally different to the usual two-centre orbital product basis used to represent charge densities in quantum chemistry, multipole analysis has become the main method employed in charge density determination and analysis over the past two decades (one notable exception being the maximum entropy method[3]).

Despite the undoubted successes of the multipole model, there are several reasons why an alternative two-centre orbital product-based representation of the charge density remains attractive for experimental charge density.

(i) Unlike multipole-fitted electron densities, which can (in principle) contain unphysical negative regions of density, densities derived from orbital or wavefunction models are constrained to be positive over all space.

(ii) It is widely recognized that correlation (in a least-squares sense) between anisotropic thermal parameters and one-centre density basis functions ultimately limits the accuracy of the multipole model, particularly if anharmonic terms are included in the thermal motion expression.[1] A two-centre expansion ought to be less susceptible to this problem.

(iii) A two-centre expansion is much more rapidly convergent (with respect to basis function angular quantum numbers), *i.e.* it spans the space more efficiently.

(iv) An experimental density expressed in a two-centre product basis of gaussian functions is very readily compared with its variationally-determined *ab initio* or density functional theory (dft) equivalent density using standard quantum chemistry codes.

In the last ten years, a few such models involving some kind of orbital or wavefunction fitting have been developed, with some applications to molecular crystals.[4,6,7] Our own previous work in this area[4] focussed on fitting an idempotent density matrix P, in a non-orthogonal basis of K functions $\{\chi_j(r)\}$ with associated overlap matrix S, *viz.*

$$\rho(r) = \sum_{i,j}^{K} P_{ij}\chi_i(r)^* \chi_j(r) \qquad (2)$$

The idempotency constraint was enforced using McWeeny's iterative algorithm.[5] Although the idempotency constraint $P S P = 2 P$ (closed-shell case) acts to considerably reduce the number of parameters to be fitted, the number of parameters in this model varies quadratically with the system size. The observations/parameters ratio therefore becomes impractically low for larger molecules and/or when larger basis sets are
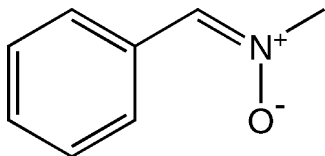
Scheme 1 A representation of (*Z*)-*N*-methyl-*C*-phenylnitrone.

utilised. Snyder and Stevens[6] implemented essentially the same scheme, but using Lagrange multipliers to enforce idempotency. Jayatilaka[7] demonstrated a "hybrid" method whereby the Hartree–Fock (HF, energy-minimizing) orbitals are additionally constrained to reproduce a set of experimental structure factors to a prescribed accuracy. This neatly circumvents the observations/parameters ratio problem as only one new parameter is required, independent of system size.

The data set used here for testing the new method has already been published as a multipole study.[8] The main chemical features of the molecule in question are a nitrone group (often drawn in charge-separated canonical form, Scheme 1) and a phenyl ring. The nitrone fragment and phenyl ring are not quite co-planar (there is a dihedral angle of about 4° between their respective planes); we will nevertheless use a qualitative σ/π classification scheme in later sections of the paper.

Computational method

We have implemented a simple scheme in which, like multipole refinement, the number of parameters required to describe the density scales linearly with the system size. The key idea is to retain the molecular orbital forms from a variational calculation (*e.g.* Hartree–Fock or dft) but to refine the occupation numbers $\{n_j\}$ of both occupied and virtual molecular orbitals $\{\phi_j(r)\}$ subject to the constraints $0 < n_j < n_{max}$ and $\sum n_j = F(000)/Z$; hence the acronym MOON (molecular orbital occupation number) refinement. The orbitals are expanded in an AO basis of contracted gaussian functions. (In this work we use Dunning's contraction of Huzinaga's double-ζ basis.[9] Split-valence gaussian basis sets are probably not suitable for this type of application since they would lead to poorer-quality core scattering factors). Eqn. (3) with $n_{max} = 2$ also represents the one-particle density in CI or perturbation methods of *ab initio* electronic structure theory..

$$\rho(r) = \sum_{j=1}^{K} n_j \phi_j(r)^* \phi_j(r) \qquad (3)$$

This density model therefore shares the advantage of our previous density matrix-based approach, in that an experimental "in-crystal" density can be, in principle, simply reproduced in an *ab initio* software package by a simple adjustment of occupation numbers. Thus the power of an *ab initio* code's property calculation section could be applied to an experimental density.

The structure factor expression associated with this model density can be written as

$$F_c(s) = \sum_{j=1}^{K} n_j \sum_{l=1}^{K} \sum_{m=1}^{K} c_{jl} c_{jm} \langle \psi_l | e^{2\pi i s \cdot r} | \psi_m \rangle \exp(2\pi i \, s \cdot r_{lm}) T_{lm}(s)$$

$$(4)$$

where c_{jl} is the (fixed) MO expansion coefficient; r_{lm} is the position vector of the lm-th basis function product; and T_{lm} is its (anisotropic) temperature factor. T_{lm} has been estimated as the average of the temperature factors for the two atoms on which a given basis function pair are centred. The basis function product Fourier transforms $\langle \psi_l | e^{2\pi i s \cdot r} | \psi_m \rangle$ are computed using the algorithm of Stewart.[10] These integrals are

calculated in an initial step and then remain fixed during refinement (they would need to be recalculated if the nuclear coordinates were altered during refinement). Nuclear coordinates $\{r_j\}$ and atomic temperature factors in this work are fixed and taken from either neutron, X-ray high-angle or X-ray multipole refinement, although in principle they could be refined simultaneously with the charge density model. The orbitals $\{\phi_j\}$ are taken from *e.g.* a Hartree–Fock or density functional theory (dft) single-point gas-phase calculation (the choice of the fixed MO basis is discussed in the Results section) on one molecule using in-crystal nuclear coordinates. The initial orbital occupation numbers, in the closed-shell case, are therefore either 2 for occupied orbitals, or 0 for virtual orbitals. We are interested in essentially two types of MOON refinement: (a) with the constraint $n_j \le 2$, so that the density is analogous to the one-particle density from a variational, full-CI procedure, and we can consider the MOs as valid spin–orbitals in the Pauli sense (maximum of two electrons per orbital); (b) no upper-limit constraint on the MO occupation numbers ($n_{max} \gg 2$). In case (a), we can view the fitting procedure as transferring electron population from the occupied to the virtual space, much as in the configuration interaction (CI) procedure. Case (b) clearly provides extra flexibility and therefore potentially a better fit, although the density has lost some of its quantum-mechanical rigour.

Eqn. (4) contains no explicit contribution for anomalous (inelastic) scattering, which arises principally from core orbitals. There are various possible approaches here.

(i) Ignore these contributions, which are in any case small if the molecule contains only light atoms. This appears to have been the approach taken in all previously published methods.[3–5]

(ii) Include the same amplitude and phase shifts used to correct atomic scattering factors[11] in the appropriate $\langle \psi_l | e^{2\pi i s \cdot r} | \psi_m \rangle$ integrals for the core orbitals.

(iii) Use a set of observed structure factors from which anomalous scattering contributions have been "removed". This requires initially phasing the $\{F_o\}$ with *e.g.* a multipole model. Such anomalous dispersion-removed structure factors are routinely required for producing Fourier synthesis maps, and are available *e.g.* from the XD program.[12]

Here we have employed option (iii), principally because (ii) logically requires that the core orbital populations should be fixed, whereas *all* orbital occupation numbers have been refined in this work. This means that the fitted density implicitly depends on the results of a multipole refinement in two ways: use of fixed, multipole-refined coordinates and temperature factors, and use of the multipole model to phase structure factors for removing anomalous dispersion. This seems justified, given that the aim of this paper is to illustrate a technique in development, rather than generate a particular density that is truly independent of multipole refinement.

The fitting procedure employed during refinement is steepest-descent optimisation based on a typical statistic

$$\chi^2 = \frac{1}{N} \sum_s w_s (|F_o(s)| - k|F_c(s)|)^2 \qquad (5)$$

where w_s is a weight based on the error in the F_o ($w_s = \sigma(F_o(s))^{-2}$ was used here) and k is the scale factor, which is re-calculated each iteration from

$$k = \frac{\sum_s w_s |F_o(s)||F_c(s)|}{\sum_s w_s |F_c(s)|^2} \qquad (6)$$

Although a unit cell electroneutrality constraint could be included in the usual matrix least-squares approach using Lagrange multipliers,[1] as implemented in the multipole refinement and analysis suite of programs XD[12] and its predecessor MOLLY,[13] the additional constraints $0 < \{n_j\} < n_{max}$ are

problematic to impose analytically. We have therefore adopted a "brute-force" approach. Elements of the gradient vector for each iteration of the optimization procedure are calculated directly from

$$\frac{\partial \chi^2}{\partial n_j} = \frac{1}{N} \sum_s \frac{-k w_s}{|F_c(s)|} (|F_0(s)| - k|F_c(s)|)$$

$$\text{Re}\left\{ F_c(s) \left(\frac{\partial F_c(s)}{\partial n_j} \right)^* + F_c(s)^* \left(\frac{\partial F_c(s)}{\partial n_j} \right) \right\} \qquad (7)$$

Using the gradient vector, the current set of occupation numbers $\{n_j\}$ are adjusted so as to lower χ^2. Shifts which would take a given orbital occupation number below 0 or above 2 (in case (a) refinement) are ignored. Electroneutrality is then restored in the following way. Two subsets of orbitals are employed: (i) the orbitals from the HF occupied set which have negative $\partial \chi^2 / \partial n_j$; (ii) the orbitals from the HF virtual set which have positive $\partial \chi^2 / \partial n_j$. If the total electron population of one molecule ($n' = \sum_j n_j$) > $F(000)/Z$, then a randomly-chosen orbital from subset (i) has its population reduced by the correct amount to restore neutrality. Similarly, the population of a randomly-chosen orbital from subset (ii) is increased to restore neutrality if $n' < F(000)/Z$. In either case, an orbital is selected whose raising or lowering of occupation number is guaranteed to further lower χ^2 because of the gradient condition. The refinement was considered converged when χ^2 was changing in several successive iterations by less than 1 part in 10^8. Gasphase single point *ab initio* calculations on (Z)-N-methyl-C-phenylnitrone at the HF/DZ and MP2/DZ(d,p) levels of theory employed Gaussian98 [14] and the multipole-refined coordinates from the published study.[8] Molecular orbital plots were produced with GaussView.[15]

The AIMPAC programs EXTREME, GRID and CONTOR[16] were employed for the topological analysis of both the MOON-refined and *ab initio* electron densities. Bond critical point topological results for the multipole-refined density were also taken from the published study.[8]

Results

Multipole refinement of the data

The details of the model employed in the XD refinement were already reported;[8] here we just repeat the key statistics. Multipole refinement on F with the 7163 reflections with $|F_o| > 4\sigma(|F_o|)$ gave a final agreement factor $R = \sum(|F_o| - k|F_c|)/\sum|F_o| = 0.040$ with a goodness-of-fit $S = [\sum w(|F_o| - k|F_c|)^2/(M-N)]^{1/2} = 1.70$. The charge density was described by a total of 168 charge density parameters (multipole coefficients and kappa expansion/contraction parameters).

MOON refinement

Integral calculation for the same 7163 structure factors used in multipole refinement and with the HF canonical molecular orbitals took approximately 24 h of CPU time on a Compaq RISC alpha workstation. The subsequent steepest-descent refinements starting from the Hartree–Fock guess ($R = 0.0547$, $S = 0.4590$, $\chi^2 = 0.2075$) took about 3 h. As described in the preceding section, two types of refinement were undertaken: (a) with $n_{max} = 2$, and (b) no upper-limit on n_{max}. In case (a) the fit statistics following refinement were $R = 0.0437$, $\chi^2 = 0.1374$ and $S = 0.3738$, respectively. With no upper-limit constraint on the occupation numbers, slightly improved statistics were obtained: $R = 0.0424$, $\chi^2 = 0.1300$ and $S = 0.3636$.

The changes in molecular orbital occupation numbers following both refinements (a) and (b) are illustrated in Fig. 1 and Fig. 2. In a double-ζ basis, the initial HF calculation for the nitrone has 36 doubly-occupied orbitals and 82 virtual orbitals, so there are 118 refined parameters in the MOON charge
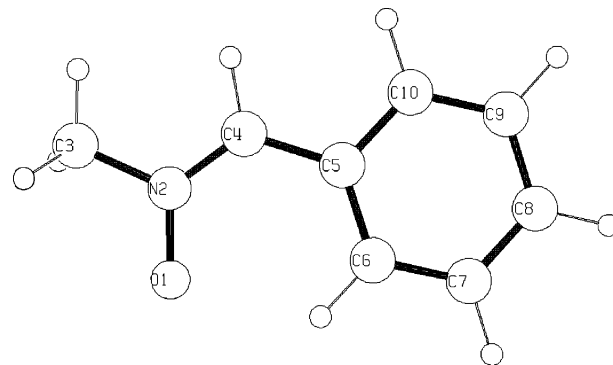

Fig. 1 Atom labelling scheme for (Z)-N-methyl-C-phenylnitrone.

density model (compared to the 168 charge density parameters in the multipole model). We remind the reader that the same fixed Hartree–Fock MOs are used in both refinements; only the occupation numbers are being adjusted.

Refinement (a). The largest occupation number for an orbital in the HF virtual set is $n_{51} = 0.655$: this orbital is a delocalized σ-type orbital with largest amplitude on the C4-N2-O1 nitrone fragment, illustrated in Fig. 3(d). Similarly, the orbital from the set 1–36 with the lowest population following refinement is the Hartree–Fock HOMO ($n_{36} = 1.451$) is illustrated in Fig. 3(b). This orbital links the phenyl ring and nitrone fragment π systems. Hence the simple picture of the experimental density is decreased π density in favour of σ density (in comparison with the HF initial density). As expected, the larger reductions in population are confined to the valence orbitals. The lowest population in any core orbital is $n_5 = 1.681$, which is a localized 1s-type orbital on the oxygen atom.

Refinement (b). The largest occupation number for an orbital in the HF virtual set is $n_{43} = 0.399$, a diffuse σ-type orbital with large amplitude on the phenyl and methyl hydrogens. This orbital is illustrated in Fig. 3(c). The orbital from the HF occupied set with the greatest depopulation following refinement is again the HOMO π-type orbital ($n_{36} = 1.289$). There is generally less depopulation of the HF core MOs compared to the constrained refinement (a). As expected, several orbitals have occupation numbers greater than 2, the largest being ($n_{15} = 2.503$), illustrated in Fig. 3(a). This is in essence the e_{1u} valence MO of benzene, with amplitude on four of the six phenyl ring atoms. Generally we observe that less density has been promoted into the HF virtual orbitals (compared to the constrained refinement) because the refinement has more flexibility in the HF occupied space.

Population analyses

The changes in Mulliken atomic charges following refinement (*i.e.* compared to the gas-phase HF wavefunction) are listed in Table 1. Also listed are the Mulliken charges from the HF/DZ wavefunction (*i.e.* the density before any refinement) and MP2-level charges computed with polarization functions added to all atoms. (Experience suggests that care should be taken when comparing Mulliken charges calculated with different basis sets; but group charges and differences between charges *e.g.* across a bond are features which are both chemically more meaningful and are less basis-set dependent.[17]) Differences of up to 0.3 electrons are seen in the atomic charges for the two types of MOON/DZ refinement; these are certainly differences which are chemically significant. In the key nitrone moiety, the constrained refinement actually suggests an almost neutral oxygen atom, although the formal charge would be −1 in the charge-separated canonical form given shown in Scheme 1.

The unconstrained refinement does result in a negatively-charged oxygen and a N–O charge difference of 0.67 electrons,
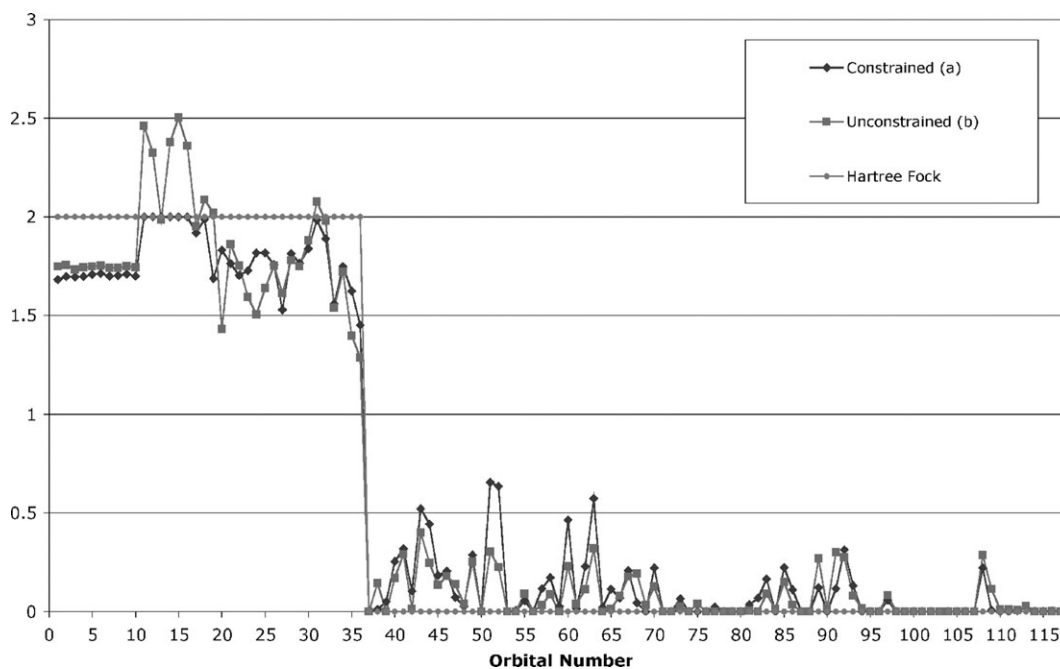
Fig. 2 Orbital occupation numbers following two types of MOON refinement with the Hartree–Fock (Dunning double-ζ) molecular orbitals.

almost the same as the 0.69 result suggested by MP2-level Mulliken charges. Both refinements suggest a much more polarized N–Me bond than in the gas-phase MP2 density, although the unconstrained refinement is closer to the MP2 result. Assuming that it is reasonable to take the MP2 density as a benchmark, then the Mulliken charges corresponding to the unconstrained refinement (b) are uniformly closer to the MP2 density than the constrained refinement (a). In particular, we note that the C5 carbon of the phenyl ring, which is bonded to the electron-withdrawing nitrone substituent, carries a similar positive charge as in the MP2-level density. Both unconstrained MOON and MP2 densities result in negative charges on all other ring carbon atoms, unlike the constrained refinement, which gives a positively-charged C10 atom.

Electron density topology

It is often convenient to characterize molecular electron densities using the scalar fields $\rho(r)$ and $\nabla^2\rho(r)$ evaluated at so-called "bond critical points"-sets of coordinates $\{r_c\}$ where $\nabla\rho = 0$ and two of the three principal curvatures of ρ are negative.[18] These data, as calculated with the programs EXTREME[16] and XDPROP[12] (for the multipole-refined density)

are presented in Table 2. Notwithstanding the effects of the crystalline environment, we might expect a good-quality experimental density to more closely resemble that of an electron-correlated gas-phase calculation. In other words, starting with the HF/DZ density (which is our initial guess) the values of $\rho_c(r)$ and $\nabla^2\rho_c(r)$ should move in the direction of the MP2/DZ(d,p) single point density which we use as a benchmark. Comparing columns (i) and (iv) in Table 2 it is evident that the MP2 density values $\{\rho_c\}$ at bond critical points are uniformly higher than the HF/DZ values, by 10–20%. Unfortunately, the density values in the bonds for the constrained refinement (ii) are uniformly *lower* than HF/DZ, whereas the values for the unconstrained refinement are, in all but one bond, higher than HF/DZ (and therefore closer to the MP2 results).

Turning now to the values $\nabla^2\rho_c$, critically we note a positive value for the N–O bond in the constrained refinement, whereas the unconstrained, multipole-refined and MP2 densities all predict a small but negative value of $\nabla^2\rho_c$. This suggests a higher ionic character for this bond in the constrained MOON refinement.[16]

$\nabla^2\rho(r)$ distributions

Fig. 4 parts (i)–(vi) depict $-\nabla^2\rho(r)$ for the two MOON refinements and the MP2-level reference density. Figs. 4 (i), (iii) and (v) are plotted in the planes of the nitrone fragment; (ii), (iv)
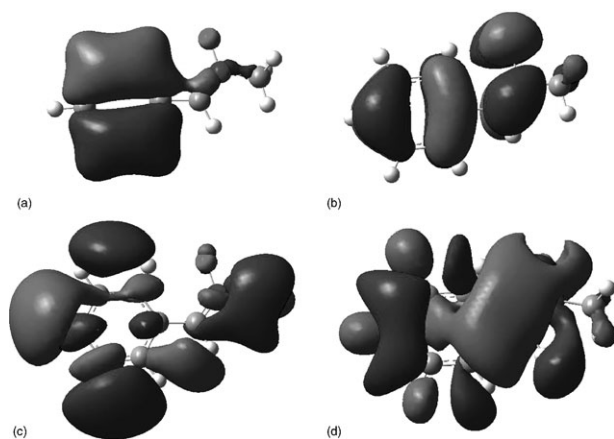

Fig. 3 HF/DZ molecular orbitals with the largest changes in occupation number after refinement: (a) orbital 15 (b) orbital 36 (c) orbital 43 (d) orbital 51.

Table 1 Mulliken charges (non-hydrogen atoms only) calculated before and after refinement; and the comparison with an electron-correlated density

	HF/DZ	MOON/DZ(a)	MOON/DZ(b)	MP2/DZ**
O1	−0.54	+0.03	−0.23	−0.44
N2	+0.04	+0.55	+0.44	+0.25
C3	−0.44	−0.88	−0.77	−0.34
C4	−0.13	−0.38	−0.23	−0.25
C5	+0.37	+0.69	+0.45	+0.36
C6	−0.32	−0.23	−0.28	−0.26
C7	−0.22	−0.71	−0.49	−0.12
C8	−0.20	−0.47	−0.17	−0.14
C9	−0.23	−0.73	−0.52	−0.13
C10	−0.34	+0.20	−0.50	−0.25

and (vi) in the mean plane of the phenyl ring. The maps in the ring plane are almost indistinguishable for the two MOON refinements, but do differ in appearance to the MP2-level density, which has generally broader regions of charge concentration enclosed by the valence shell of charge concentration (VSCC). In the nitrone fragment plane, the constrained MOON refinement shows a weak concentration of density which lies between the two out-of-plane methyl hydrogens; this feature is not present in the unconstrained refinement, nor in the MP2 density. It is also striking that the oxygen atom is a completely detached "island" of charge concentration in the constrained refinement, whereas the unconstrained MOON and MP2 densities predict a "pinched" VSCC which is nevertheless jointed continuously with the VSCC around nitrogen. This is consistent with the positive value of $\nabla^2\rho_c$ reported in Table 2 for the constrained refinement, which is out of line with all the other densities.

Discussion

From the results presented in the preceding section, it would appear that the density obtained from the unconstrained MOON refinement is more similar to an electron-correlated density than the constrained refinement, and therefore presumably of superior quality. We tentatively interpret this finding as follows. The Hartree–Fock virtual orbitals alone do not provide a sufficiently flexible basis of additional functions for describing the experimental density. It appears that they do not contain enough functions which would allow density to increase in the bonding regions (i.e. they are mostly of a diffuse nature). The fitting procedure in the "constrained" MOON refinement involves occupying these virtual orbitals at the expense of the population in the HF occupied orbitals. In addition to allowing the virtual orbitals to become occupied, the unconstrained procedure permits the population of the HF occupied orbitals to rise above two; indeed this did take place in eight of the 36 orbitals. The extra flexibility in this unconstrained model appears to be important. For the moment this remains a tentative conclusion, since (i) it is based on one example, (ii) it is based on the HF orbitals only, i.e. perhaps a different choice of fixed MO basis (Kohn–Sham, natural orbitals, Boys-localized orbitals etc.) might prove to have more of the required flexibility in the virtual space, (iii) basis set deficiency (our current Fortran code only supports sp basis sets) might also be playing a role here. Since the unconstrained MOON refinement permits orbital occupation numbers greater than two, some of the quantum-mechanical rigour of the corresponding electron density is lost, because it can no longer be identified with a linear combination of Slater determinants of doubly-occupied spin orbitals. However, this is no different to other methods such as multipole refinement where the fitted functions have no particular quantum-mechanical significance; they just represent a convenient and compact basis for describing the density. Moreover, the unconstrained MOON
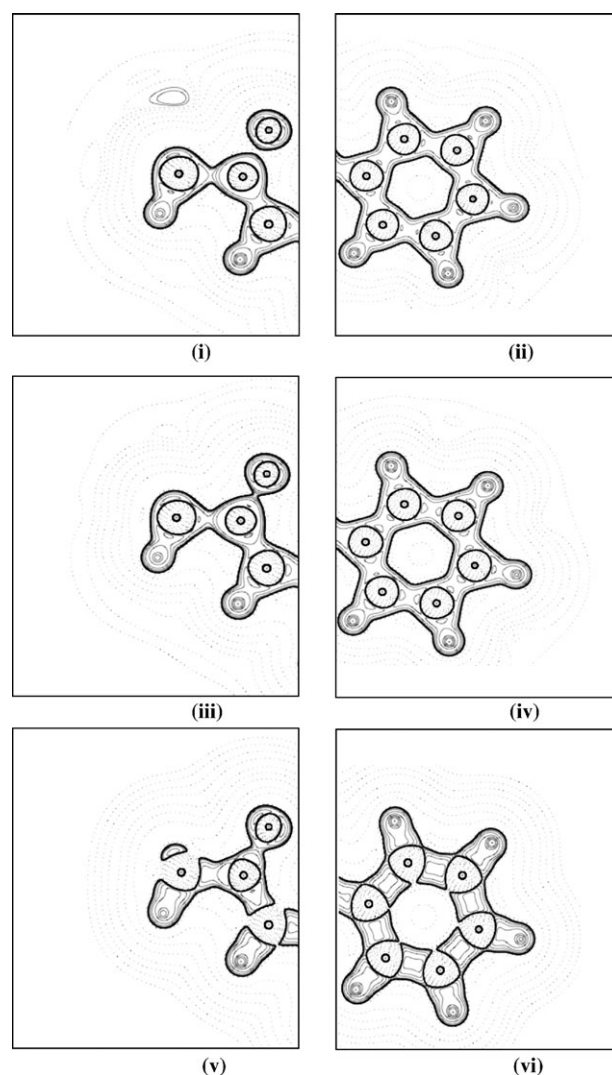

Fig. 4 Contour maps of $-\nabla^2\rho(\boldsymbol{r})$, solid lines representing positive values. (i) and (ii) MOON/DZ constrained; (iii) and (iv) MOON/DZ unconstrained; (v) and (vi) MP2/DZ(d,p). All densities computed using the X-ray refined nuclear coordinates from ref. 7. Contours at $(2, 4, 8 \times 10^n, n = -3, -2, -1, 0, 1, 2)$ e Å^{-5}.

refinement is still more quantum mechanically-rigorous than multipole refinement because it retains the essential constraint of a positive density at all points in space.

Whether the choice of fixed MO basis for the MOON refinements strongly influences the results obtained is clearly an important question. In order to explore this, two additional sets of MOON refinements were carried out: (i) with Kohn–Sham orbitals obtained from a B3LYP/DZ single point, (ii) with natural orbitals obtained by diagonalizing the HF/DZ

Table 2 Density topological properties from various models. All densities computed at the multipole-refined nuclear coordinates

(3,−1) CP	HF/DZ		MOON/DZ constrained		MOON/DZ unconstrained		MP2/DZ(d,p)		Multipole refined	
	ρ_c/e Å^{-3}	$\nabla^2\rho_c/e$ Å^{-5}	ρ_c/e Å^{-3}	$\nabla^2\rho_c/e$ Å^{-5}	ρ_c/e Å^{-3}	$\nabla^2\rho_c/e$ Å^{-5}	ρ_c/e Å^{-3}	$\nabla^2\rho_c/e$ Å^{-5}	ρ_c/e Å^{-3}	$\nabla^2\rho_c/e$ Å^{-5}
N2–O1	2.53	−5.5	2.44	+2.4	2.72	−5.6	2.76	−12.9	3.00	−3.3
N2–C3	1.55	−9.1	1.52	−6.7	1.56	−8.0	1.73	−15.5	1.81	−11.6
N2–C4	2.13	−8.5	2.12	−14.2	2.21	−14.6	2.25	−5.2	2.56	−26.8
C4–C5	1.77	−16.0	1.68	−12.1	1.77	−14.4	1.89	−20.0	1.94	−13.3
C5–C10	1.93	−18.6	1.84	−15.7	1.94	−17.6	2.06	−22.8	2.15	−16.5
C10–C9	1.98	−19.6	1.90	−15.2	1.98	−17.5	2.11	−23.9	2.23	−17.6
C9–C8	1.96	−19.2	1.88	−14.3	1.95	−16.6	2.09	−23.6	2.21	−17.4
C8–C7	1.97	−19.4	1.91	−16.8	1.96	−18.0	2.10	−23.9	2.22	−18.7
C7–C6	1.97	−19.5	1.89	−14.7	1.98	−17.3	2.10	−23.7	2.20	−17.3
C6–C5	1.92	−18.5	1.85	−14.8	1.94	−16.7	2.04	−22.5	2.14	−16.0

Table 3 ρ_c values for selected bonds with various densities ($e\ \text{Å}^{-3}$)

	HF/DZ	MOON/ DZ(HF)	B3LYP/ DZ	MOON/ B3LYP(DZ)	DZNO	MOON/ DZNO
N2–O1	2.53	2.72	2.56	2.83	2.56	2.87
N2–C3	1.55	1.56	1.60	1.65	1.59	1.65
N2–C4	2.13	2.21	2.13	2.25	2.13	2.26
C4–C5	1.77	1.77	1.77	1.80	1.77	1.76

density matrix-referred to as DZNO in Table 3. (Natural orbitals tend to be more localized than their canonical HF or Kohn–Sham counterparts.) In both cases the refinements were carried out with no upper-limit on the occupation numbers, so they were equivalent to the type (b) refinement with the HF /DZ orbitals. The key statistics obtained with the new refinements were: (i) $R = 0.0422$, $\chi^2 = 0.1287$ and (ii) $R = 0.0431$; $\chi^2 = 0.1327$). Comparing this with the refinement using HF/ DZ orbitals ($R = 0.0424$, $\chi^2 = 0.1300$) we observe that the Kohn–Sham orbital refinement gives a slightly better fit, while the natural orbitals give a marginally worse result. In order to gauge how different are the three MOON-refined densities using different MO bases, Table 3 compares bond critical point densities for the main bonds in the nitrone moiety. In each case, the ρ_c values are given both before and after MOON refinement. The refinement process uniformly increases the bond critical point densities in almost all bonds for all three MO bases. In two of the bonds (N2–O1 and N2–C3) there are notable differences in the final refined values (e.g. the Kohn–Sham and natural orbital bases give higher values for these two bonds by 4–6%). This indicates that the various MO bases derived from the Dunning/Huzinaga DZ atomic orbital (AO) basis are not "saturated" in the sense that they do not contain enough functions to completely describe the experimental density. However, encouragingly these differences are small enough to suggest that convergence in the experimental density (regardless of MO basis) could be attained by increasing the size of the AO basis to include e.g. polarization functions and extra sp shell of functions.

Conclusions

In summary: these preliminary results demonstrate that the molecular orbital occupation number (MOON) refinement is a viable orbital-based method for analysing experimental X-ray elastic scattering data, which does not require an excessive number of parameters to describe the density distribution. It should therefore be applicable to much larger molecules than those studied in previous applications of orbital/wavefunction-based charge density analysis. Reassuringly there are no great swings in the occupation number (compared to the gas-phase) for any orbital, and only small reductions in population for core orbitals. The orbital-based nature of the model lends the fitted density to chemical interpretation, e.g. in the nitrone a

depopulation of the ring-nitrone π system is observed, with a concomitant increase in σ density. (Using the fitted orbital populations it would be possible to carry out more detailed types of orbital analysis, for example computing Meyer bond orders.[19]) In crystals where strong intermolecular interactions are present, it might also prove possible to rationalize the changes in specific orbital populations (i.e. compared to the gas-phase) by considering the crystalline environment. The quality of the fitted experimental density for the nitrone has been verified by comparison with an electron-correlated density calculated using the same nuclear coordinates. These results suggested that the MOON-refined density is comparable in quality to the multipole-refined density, despite the lower number of parameters required.

Since several related models for wavefunction-based charge density analysis have now been presented in the literature, it seems pertinent to compare their various strengths and weaknesses. We summarize the main points in Table 4.

The earlier Howard–Huke and Stevens–Snyder models are in most respects identical, differing only in the details of the idempotency constraint. Both are only applicable to small molecules due to the quadratic rate of increase in the required parameters of a density matrix. However they have the attractive feature of (in principle) generating the exact, electron-correlated density with a single determinant, i.e. they give the Kohn–Sham orbitals, without recourse to an energy-minimization criterion. The Jayatilaka model in effect minimizes a mixed statistic which is a linear combination of the Hartree–Fock (gas-phase) energy and the same χ^2 statistic used in this work. It has the remarkable attribute of requiring just one extra parameter for any system size (that parameter determining the relative importance of minimizing the energy versus reproducing experimental structure factors). It is, however, inextricably married to the HF mean field electronic structure model, which could be disadvantageous in systems where electron correlation effects are likely to dominate.

The molecular orbital occupation number refinement model proposed here has linear scaling; relative ease of chemical interpretation; and is capable of returning the "exact" electron density. A key result of a refinement is an experimental wavefunction (wfn) input file which is compatible with all the major computer programs for topological and atoms-in-molecules analysis of electron densities. The lack of idempotency means that property computation is restricted to one-electron properties (e.g. no energies). Arguably this is not a serious restriction, since the same restriction applies to densities obtained by the popular multipole refinement. Also the fixed orbital forms means that it will never be quite as flexible as a full refinement of the (idempotent) density matrix.

In short, it would appear that no one orbital-based model stands out as the clear method-of-choice; all have subtly different things to offer. Weak points with all of the orbital models presented to date include (i) the accuracy of core scattering when using a gaussian basis, and (ii) the treatment

Table 4 Comparison of orbital-based models for charge density analysis

	Idempotent, so single-determinant wavefunction is available	Scaling of parameters with respect to system size	Additional, energy-minimizing criterion used?	Ease of interpretation of the fitted wavefunction/orbitals
Howard, Huke et al., 1994 (ref. 4)	Yes	$\approx N^2$	No	Complicated to interpret
Jayatilaka, 1998 (ref. 7)	Yes	Only one parameter needed	Yes	Complicated to interpret
Snyder and Stevens, 1999 (ref. 6)	Yes	$\approx N^2$	No	Complicated to interpret
Hibbs, Howard, Huke and Waller (this work)	No	$\approx N$	No	Fixed orbital forms, so changes in occupation number easily interpreted

of anomalous scattering. Further work is needed to establish the best way forward with these problems.

Acknowledgements

References

1 P. Coppens, *X-Ray Charge Densities And Chemical Bonding*, Oxford University Press, Oxford, 1997.
2 T. S. Koritsanszky and P. Coppens, *Chem. Rev.*, 2001, **101**, 1583.
3 S. Aoyagi, Y. Kuroiwa, A. Sawada, E. Nishibori, M. Takata, M. Sakata, H. Tanaka and J. Harada, *Ferroelectrics*, 2002, 1151.
4 S. T. Howard, J. P. Huke, C. S. Frampton and P. R. Mallinson, *Phys. Rev. B*, 1994, **49**, 7124.
5 R. McWeeny, *Rev. Mod. Phys.*, 1960, **32**, 335.
6 J. A. Snyder and E. D. Stevens, *Chem. Phys. Lett.*, 1999, **313**, 293.
7 D. Jayatilaka, *Phys. Rev. Lett.*, 1998, **80**, 798.
8 D. E. Hibbs, J. R. Hanrahan, M. B. Hursthouse, D. W. Knight, J. Overgaard, P. Turner, R. O. Piltz and M. P. Waller, *Org. Biomol. Chem.*, 2003, **1**, 1034.
9 T. H. Dunning, *J. Chem. Phys.*, 1970, **53**, 2823.
10 R. F. Stewart, *J. Chem. Phys.*, 1969, **51**, 4569.
11 L. Kissel, B. Zhou, S. C. Roy, S. K. Sen Gupta and R. H. Pratt, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 1995, **51**, 271.
12 T. Koritsanszky, P. R. Mallinson, S. T. Howard, A. Volkov, P. Macchi, Z. Su, C. Gatti, T. Richter, L. J. Farrugia and N. K. Hansen, *XD – A Computer Program Package for Multipole Refinement and Analysis of Electron Densities from Diffraction Data*, Free University of Berlin, Berlin, Manual Version 12, 2003.
13 N. K. Hansen and P. Coppens, *Acta. Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Cryst.*, 1978, **34**, 909.
14 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Ragha-vachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *GAUSSIAN 98 (Revision A.6)*, Gaussian, Inc., Pittsburgh, PA, 1998.
15 A. E. Frisch, R. D. Dennington II, T. A. Kieth, A. B. Nielsen and A. J. Holder, *GaussView Reference Manual*, Gaussian, Inc., Pittsburgh, PA, 2003.
16 F. W. Biegler-Koenig, R. F. W. Bader and T. H. Tang, *J. Comput. Chem.*, 1982, **3**, 317.
17 A. Hinchliffe, *Computational Quantum Chemistry*, J. Wiley and Sons, Chichester, 1988.
18 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, 1990.
19 I. Mayer, *Chem. Phys. Lett.*, 1983, **97**, 270.