Acta Cryst. (1997). B53, 553–564

Retrieval of Structure-Factor Phases in Non-centrosymmetric Space Groups. Model Studies Using Multipole Refinements

MARK A. SPACKMAN* AND PATRICK G. BYROM

Department of Chemistry, University of New England, Armidale NSW 2351, Australia.
E-mail: mspackma@metz.une.edu.au

(Received 23 July 1996; accepted 9 December 1996)

Abstract

Structure factors with known magnitude and phase are constructed for seven different molecular crystals in non-centrosymmetric space groups and used to calibrate the retrieval of structure-factor phase information with several different scattering factor models. Monopole models are confirmed to be inadequate, yielding poor estimates of phases and underestimating the Fourier deformation electron density by as much as 50%. Multipole models are generally successful, with the conspicuous exception of hexamethylenetetramine (HMT). The results confirm the known difficulty of phasing structure-factor magnitudes for HMT, but reveal a broad spectrum of behaviour of phase retrieval in various non-centrosymmetric space groups, depending partly on the number of reflections with restricted phases. For several systems (urea, hydrogen peroxide, L-alanine and borazine) the results confirm that multipole refinements of scattering factor models against experimental data are capable of yielding highly accurate phases and hence reliable electron distributions. For the exceptional case of HMT a simple eigenvalue filtering technique enables retrieval of phases and electron densities with an accuracy comparable to the best of the other non-centrosymmetric systems studied.

1. Introduction

In the charge density analysis of X-ray diffraction data for non-centrosymmetric space groups it is widely recognized (but not always!) that only structure-factor magnitudes can be obtained directly from the experiment; the phases are model dependent (see, for example, Coppens, 1982; Blessing & Lecomte, 1991). In order to determine the reflection phases it is necessary to fit an appropriate structure-factor model to the amplitude data and then infer the values of the phases. Although the generally accepted criterion for a successful fit to the amplitudes is the lowest weighted sum of the squared residuals between the observed and calculated structure-factor magnitudes, this does not guarantee that the phases calculated from the model will be adequate to produce a chemically significant electrostatic property. While phases from centrosymmetric space groups are usually well determined (except for small reflections), the same confidence cannot always be placed on phases deduced from non-centrosymmetric space groups.

Some time ago Coppens (1974) and Feil (1977) warned of the hazards in extracting electrostatic properties from X-ray diffraction data derived from non-centrosymmetric crystals, but since then there have been many, apparently successful, charge density studies reported on non-centrosymmetric crystalline materials. However, very few of these studies have analysed the changes in phase arising from various models employed to determine phases. Exceptions are the pioneering work on sucrose (Hanson, Sieker & Jensen, 1973), urea and thiourea (Mullen & Scheringer, 1978; Mullen, 1980, 1982), $LiHCOO.H_2O$ (Thomas, 1978) and H_2O_2 (Savariault & Lehmann, 1980), as well as the more recent work by Lecomte and co-workers on peptides and related molecules (Lecomte, 1991; Souhassou, Lecomte, Blessing, Aubry, Rohmer, Wiest, Bénard & Marraud, 1991; Lecomte, Ghermani, Pichon-Pesme & Souhassou, 1992; Souhassou, Lecomte, Ghermani, Rohmer, Wiest, Bénard & Blessing, 1992; Wiest, Pichon-Pesme, Bénard & Lecomte, 1994). Most recently, El Haouzi, Hansen, Le Hénaff & Protas (1996) discussed this aspect of the phase problem in some detail, with specific reference to GaAs and LiB_3O_5. A variety of structure-factor models have been used in many of these studies. The atomic positional and thermal parameters have often been fixed at values derived from neutron diffraction or refined using the X-ray data. Standard atomic form factors have been used and also a variety of multipole expansions of the atomic charge density, and even a bond charge model. While phase angles derived from different models have been compared, no studies have calibrated the extraction of phase angles against a realistic set of theoretically derived phases. Generally, more flexible models have been considered to produce more reliable estimates of the phases.

The present study attempts to examine this problem through the use of model structure factors with known

phases. These are generated for several non-centrosymmetric space groups that incorporate electron densities for simple molecules derived from accurate Hartree–Fock wavefunctions. The sets of structure factors are then used as 'observed' structure factors in refinements using several different multipole models and the calculated phases compared with the 'observed' phases. In addition, deformation density maps are plotted, phasing the 'observed' amplitudes with the results of the various refinements.

Our focus throughout is on organic molecular crystals rather than high-symmetry inorganics and our results should complement those of El Haouzi *et al.* (1996). We note that experimental determination of phases is an active area of research (see, for example, Shen & Colella, 1986; Hümmer, Weckert & Bondza, 1990; Spence, 1993), although the only useful results (for the purposes of charge density analysis) appear to be measurements of the phase of the 002 reflection (with estimated error 0.07°) in CdS (Zuo, Spence & Hoier, 1989) and the 002 reflection in BeO [with estimated error 0.52° (Zuo, Spence, Downs & Mayer, 1993)]; for organics, triplet phases can be determined with an accuracy of ~45° (Hümmer *et al.*, 1990).

2. The importance of non-centrosymmetric space groups in charge density analysis

What fraction of charge density studies involves non-centrosymmetric space groups? In a recent review of charge density studies published between 1992 and 1994 (Spackman & Brown, 1994), the ratios of non-centrosymmetric to centrosymmetric studies in various categories were found to be:

inorganic materials:	8/50 = 16%
coordination compounds:	6/16 = 38%
molecular crystals (excluding biomolecules)	10/42 = 24%
biomolecules:	5/11 = 56%.

Moreover, the space groups $P2_12_12_1$ and $P2_1$ between them account for either 29.7 (Donohue, 1985) or 18.3% (Mighell, Himes & Rodgers, 1983) of the known organic molecular crystal structures, depending on how the counting is done (see Srinivasan, 1991). These statistics, coupled with the preponderance of non-centrosymmetric space groups for biomolecules and the increasing interest in the properties of crystalline non-linear optical materials (Howard, Hursthouse, Lehmann, Mallinson & Frampton, 1992; Radaev, Maximov, Simonov, Andreev & D'yakov, 1992; Fkyerat, Guelzim, Baert, Paulus, Heger, Zyss & Périgaud, 1995), for which the space group must be non-centrosymmetric (Nicoud & Twieg, 1987), suggest that increasing numbers of such studies will be performed in the future.

From the existing studies of model phase dependence cited above it has been firmly established that the use of an inadequate phasing model (*e.g.* spherical atoms) typically leads to an underestimate in peak heights in Fourier deformation electron density maps, and by as much as 50% (Savariault & Lehmann, 1980), although this is clearly space-group-dependent because of the varying contribution from restricted phases (see below). Although most current charge density studies performed on non-centrosymmetric molecular cystals seek more than a Fourier deformation density, a number of recent analyses have presented electron density maps as the sole basis for any chemical discussion (see, for example, Belaj, 1992; Radaev *et al.*, 1992; Boese, Maulitz & Stellberg, 1994; Antipin, Chernega, Lysenko, Struchkov & Nixon, 1995), sometimes without acknowledging the existence of a phase problem *per se*.

3. Description of the problem and outline of procedure

Figure 1 summarizes our attack on the problem and the various terms are defined as

$F_t(\mathbf{H}) = |F_t(\mathbf{H})| \exp\left(i\varphi_t(\mathbf{H})\right)$, the 'true' structure factor and phase;

$F_m(\mathbf{H}) = |F_m(\mathbf{H})| \exp\left(i\varphi_m(\mathbf{H})\right)$, the model structure factor and phase;

$\Delta\varphi = \varphi_t - \varphi_m$, the phase angle difference (mod 2π).

The starting point of the study is the generation of accurate Hartree–Fock wavefunctions using a polarized double-zeta basis set with nuclear coordinates obtained from known crystal structures. From the superposition

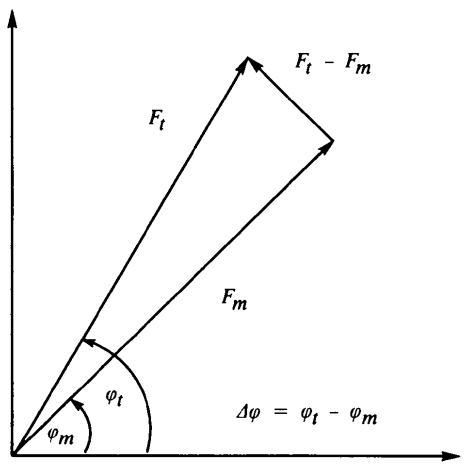

Fig. 1. A statement of the phase problem as pursued in this study. F_t is the 'true' structure factor derived from a superposition of molecular electron densities in the unit cell, F_m the model structure factor (either spherical atom or multipole model), and φ_t and φ_m their respective phases.

Table 1. *The molecular crystals used in the study, their space group, the number of reflections generated for each* $((\sin\theta/\lambda)_{max} = 1.0\,\text{Å}^{-1})$, *the number and percentage of reflections with restricted phases and the reference for the crystal data*

Molecule	Space group	$N_{reflections}$	$N_{restricted}$	Reference
Acetamide	$R3c$	1382	11 (<1%)	Jeffrey *et al.* (1980)
Hydrogen peroxide	$P4_12_12$	350	176 (50%)	Savariault & Lehmann (1980)
Urea	$P\bar{4}2_1m$	403	133 (33%)	Swaminathan, Craven & McMullan (1984)
Borazine	$P4_32_12$	1252	490 (39%)	Boese *et al.* (1994)
Cyclopropane	$Cmc2_1$	716	89 (12%)	Nijveldt & Vos (1988)
Hexamethylene-tetramine (HMT)	$I\bar{4}3m$	164	44 (27%)	Kampermann *et al.* (1994).
L-Alanine	$P2_12_12_1$	2033	535 (26%)	Destro, Marsh & Bianchi (1988)

of these molecules in the unit cell, two sets of structure factors were generated; one set ('dynamic') incorporates experimental thermal motion, while the other set ('static') does not. Full details of the method have been given in a previous paper (Spackman & Byrom, 1996) and will not be repeated here. These sets of structure factors have known magnitudes and phases and, therefore, enable us to determine how well a number of structure factor models retrieve both the phases *and* magnitudes of $F_t(\mathbf{H})$ and, just as importantly, the errors arising from approximating φ_t by φ_m in a Fourier map of $\Delta\rho$. We emphasize that this is distinctly different from all previous discussions of phase errors, where the phase difference has been either that between two spherical atom models [*e.g.* using X-ray (X) and neutron (N) position and thermal parameters] or between a multipole model and a spherical atom model.

Table 1 lists details for the seven non-centrosymmetric molecular crystals studied, in particular the number ($N_{restricted}$) and proportion of reflections with phases restricted by the space group's symmetry [details of the classes of reflections involved in individual cases can be obtained from Table 3.1 of Giacovazzo (1992)]. The role of restricted phases is an interesting one, which has been discussed previously by Souhassou *et al.* (1991), and Table 1 highlights the differences for these space groups. Our choice of crystals ranges from acetamide, where less than 1% of the reflections have restricted phases, to hydrogen peroxide, for which half of the reflections need not be phased by the model. As we demonstrate below, these proportions have a direct bearing on the degree of success in retrieving the 'true' phases.

For each molecular crystal, $|F_t(\mathbf{H})|$ were then fitted with four different models using *VALRAY* (Stewart & Spackman, 1983), with unit weights applied to each of the structure factors. All refinements were based on $|F|$, with fixed positional parameters, and thermal motion parameters either fixed at zero (static data) or the appropriate experimental X-ray or neutron values used to generate the model data (dynamic data). Four scattering factor models were refined against each data set:

3.1. *IAM: Independent atom model*

Spherical atomic scattering factors (Clementi & Roetti, 1974) were used with an overall scale factor. Since these refinements did not involve variation of atomic position and thermal parameters, this is rather different from (and less flexible than) the spherical atom models described previously by others (*e.g.* Thomas, 1978).

3.2. *A: Monopoles only*

A localized core (Clementi & Roetti, 1974; Stewart, 1980) was used for non-hydrogen atoms, with all core populations constrained to be equal. Localized valence functions were used to describe the valence shell of all atoms (single exponential in the case of H) and a population parameter and radial scale factor was refined for each of these valence functions, subject to the constraint that all atoms of the same type have the same scale factor (but different populations). This model is basically the same as the κ-refinement model (Coppens, Guru Row, Leung, Stevens, Becker & Yang, 1979).

3.3. *B: Multipole model with fixed exponents*

As for *A* plus higher multipoles on each atom (dipoles, quadrupoles and octopoles on B, C, N, O; dipoles and quadrupoles on H) with single exponential radial functions, $r^n \exp(-\alpha r)$, the radial parameters fixed at the standard molecular (SM) values (Hehre, Stewart & Pople, 1969; Hehre, Ditchfield, Stewart & Pople, 1970). For H we chose $n = 0, 1, 2$ for monopoles, dipoles and quadrupoles, respectively, and for all other atoms $n = 2, 2, 3$ for dipoles, quadrupoles and octopoles, respectively. The same coordinate system was used for all atomic multipole functions and all functions allowed by the site symmetry of the atom were included (*i.e.* approximate molecular symmetry was not imposed on the pseudo-atom model).*

3.4. *C: Multipole model with optimized exponents*

As for *B* with radial parameters optimized, but constrained such that all higher-multipole radial

*There appear to be some errors in the relationships between multipole populations for HMT listed by Terpstra, Craven & Stewart (1993) in their Table 2. Straightforward group theory analysis provides: C (2*mm*) $q_5 = -0.6052q_1$, $o_5 = -0.7915o_1$; N (3*m*) $o_5 = o_6 = -0.7915o_1$, $o_7 = 1.3829o_1$; H (*m*) $q_5 = -0.6052q_1$. All other relationships are correct.

parameters on each particular atom type are equal. For H atoms, monopole exponents/scale factors were optimized independently of higher multipole exponents.

In order to assess the retrieval of the true structure-factor phases several statistics were generated for each structure-factor model:

(i) mean (signed) phase angle error (reported in °):

$$\langle \Delta\varphi \rangle = \Sigma \Delta\varphi / N_{\text{unrestricted}};$$

(ii) root-mean-square (r.m.s.) phase angle error (reported in °)

$$rms(\Delta\varphi) = [\Sigma(\Delta\varphi)^2 / N_{\text{unrestricted}}]^{1/2};$$

(iii) mean (signed) arc-length error (electrons, with $\Delta\varphi$ in radians)

$$\langle F_t - F_M \rangle \simeq \langle |F_t| \Delta\varphi \rangle = \Sigma |F_t| \Delta\varphi / N_{\text{unrestricted}};$$

(iv) r.m.s. arc length error (electrons, with $\Delta\varphi$ in radians)

$$\text{r.m.s.}(|F_t|\Delta\varphi) = [\Sigma |F_t|^2 (\Delta\varphi)^2 / N_{\text{unrestricted}}]^{1/2};$$

(v) approximate r.m.s. error in the Fourier electron density (e Å^{-3}, with $\Delta\varphi$ in radians)

$$\text{r.m.s.}(\delta\rho) = V^{-1}[\Sigma \text{mult} \times |F_t|^2 (\Delta\varphi)^2]^{1/2}.$$

The summations in (i)–(v) are over only the symmetry-unique *hkl* and hence just the unrestricted reflections, while that in (v) involves all symmetry equivalents (*i.e.* it includes the multiplicities of the unrestricted reflections).*

Both $\langle \Delta\varphi \rangle$ and r.m.s.($\Delta\varphi$) are straightforward quantities, but the other statistics deserve further comment. The arc-length error $|F_t|\Delta\varphi$ arises from the Fourier residual electron density, in this case the difference between phasing $|F_t|$ with the true phases and the model phases

$$\begin{aligned}
\delta\rho(\mathbf{r}) &= V^{-1} \sum_{\text{sym}} \left(|F_t| \exp(i\varphi_t) \right. \\
&\quad \left. - |F_t| \exp(i\varphi_m) \right) \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}) \\
&= V^{-1} \sum_{\text{sym}} |F_t| \{ \exp(i\varphi_t) - \exp(i\varphi_m) \} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}) \\
&\simeq V^{-1} \sum_{\text{sym}} |F_t| \Delta\varphi \exp\left(i\left\{\varphi_t + \tfrac{\pi}{2}\right\}\right) \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}),
\end{aligned}$$

where the approximation in the last line is valid for small $\Delta\varphi$ (see Appendix B in Souhassou *et al.*, 1991). The expression (v) above for the approximate r.m.s. error in the electron density is derived readily from this expression using Parseval's theorem (see *e.g.* Coppens & Hansen, 1977).

* As commonly observed for centrosymmetric structures, in several refinements some reflections with very small $|F|$ exhibited phase changes of $\pm\pi$. The *worst* instances were borazine:static (*IAM*; *A*) and L-alanine:dynamic (*A*) where five such reflections behaved in this manner. To avoid undue bias in the statistics, all phase changes of $\pm\pi$ were omitted from these sums.

4. Results and discussion

Histograms summarizing the statistics (i)–(v) for models *A*, *B* and *C* refined against static and dynamic data for all molecules are given in Figs. 2–6. For clarity

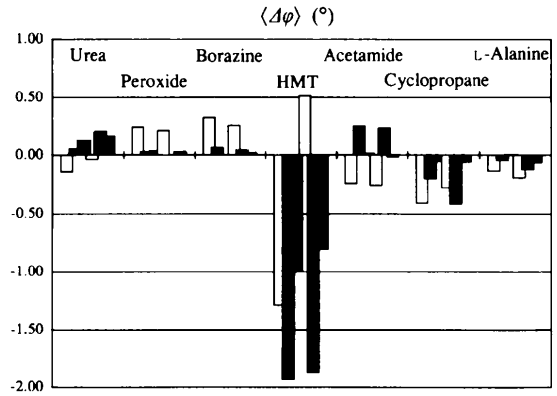

Fig. 2. Mean phase angle error, $\langle \Delta\varphi \rangle$ (°), for six models. For each molecular crystal the bars represent, from left to right: *A*: static (white), *B*: static (grey), *C*: static (black), *A*: dynamic (white), *B*: dynamic (grey) and *C*: dynamic (black). Results for the *IAM* model have been omitted, but are similar to those for model *A*.

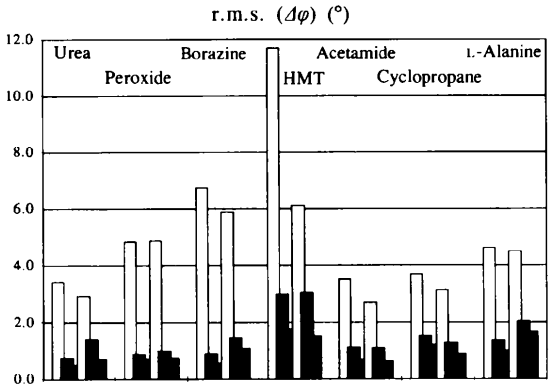

Fig. 3. R.m.s. phase angle error, r.m.s.($\Delta\varphi$) (°), for six models. The key to the bars is given in the caption to Fig. 2.

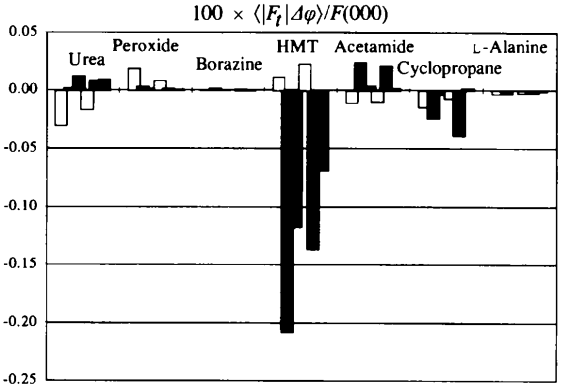

Fig. 4. Mean arc-length error, $\langle |F_t| \Delta\varphi \rangle$ [normalized to $F(000) = 100$], for six models. The key to the bars is given in the caption to Fig. 2.

we have omitted all *IAM* results from these figures as they closely parallel those for model *A* (κ refinement) reported in the figures. Also, recognizing that $F(000)$ ranges from 64 (urea) to 576 (acetamide) for our sample of molecular crystals, we have normalized results for $\langle|F_t|\Delta\varphi\rangle$ and r.m.s.$(|F_t|\Delta\varphi)$ (Figs. 4 and 5) to an $F(000)$ value of 100.

Mean and root-mean-square (r.m.s.) phase angle errors are summarized in Figs. 2 and 3, respectively. For an unbiased outcome we expect a mean phase error near zero and this is confirmed for most models and systems with the striking exception of HMT. For all other systems, $|\langle\Delta\varphi\rangle| \leq 0.50°$ for all scattering factor models and, generally, but not always, multipole models perform better than a monopole model. In contrast, for HMT the mean error is often several degrees and multipole models can actually provide a mean phase error far worse than that from a spherical atom model. Perhaps a better measure of the overall retrieval of phases is the r.m.s. phase error (Fig. 3), since a model yielding large positive and negative phase errors may accidentally result in $\langle\Delta\varphi\rangle$ near zero. From Fig. 3 it is again obvious that HMT is outstanding, but it is also much clearer here that model *A* (typical r.m.s. error of 3.0° or more) is always inferior to either of the multipole models *B* or *C* (typical r.m.s. errors near 1.0°) and, moreover, model *C* (optimized exponents) always provides a better retrieval of phases than model *B* (fixed exponents).

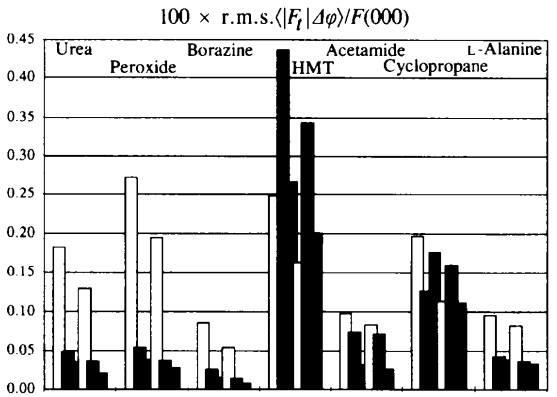

Fig. 5. R.m.s. arc-length error, r.m.s.$(|F_t|\Delta\varphi)$ [normalized to $F(000) = 100$], for six models. The key to the bars is given in the caption to Fig. 2.

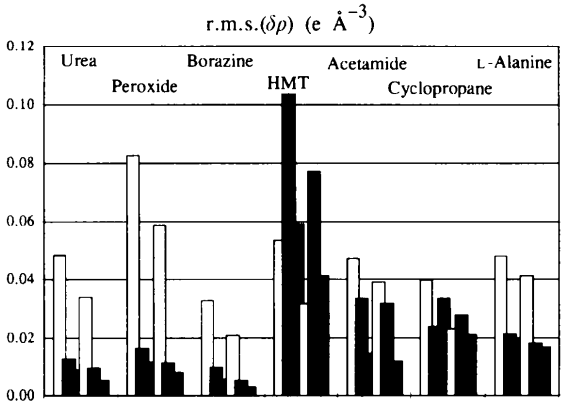

Fig. 6. Estimated r.m.s. error in the Fourier electron density, r.m.s.$(\delta\rho)$ (in e Å^{-3}), for six models. The key to the bars is given in the caption to Fig. 2.

The raw phase angle errors do not quite tell the whole story, however, since it is really the combination of the phase error, the magnitude of the structure factor and its location in reciprocal space which are important; a large phase error coupled with a small $|F_t|$ will often be of little consequence. For this reason, several recent experimental charge density studies report plots of an arc-length error, $|F|\Delta\varphi$, *versus* $\sin\theta/\lambda$ or $|F|$, where $\Delta\varphi$ is the phase angle difference between multipole and spherical atom models (Souhassou *et al.*, 1991; Fkyerat *et al.*, 1995; Hamzaoui, Baert & Wojcik, 1996). In a similar vein, in Figs. 4 and 5 we summarize mean and r.m.s. arc-length errors for the systems and models studied. In both cases the vertical axis is in units of electrons [the results scaled to $F(000) = 100$ electrons] and hence can be readily interpreted in the context of either experimental errors or the typical magnitude of the deformation density contribution to structure factors. We see that the normalized mean arc-length error (Fig. 4) is within 0.05 electrons of zero for all systems and models, with the exception of HMT again. Curiously, models *B* and *C* are seen to perform significantly worse than model *A* for HMT and also model *B* is not as good as model *C* for HMT, acetamide and cyclopropane. Turning to $100 \times$ r.m.s.$(|F_t|\Delta\varphi)/F(000)$ (Fig. 5) HMT again stands out, but cyclopropane is also a notable outlier. For urea, peroxide, borazine and L-alanine, models *B* and *C* are clearly superior to *A*, and *C* slightly better than *B* in all cases. However, for HMT both multipole models are inferior to the κ-refinement model, for cyclopropane the multipole models represent at best a slight improvement over model *A*, while for acetamide model *B* is marginally better than *A*, and model *C* dramatically better than *B* in turn. We conclude that in well behaved cases (*e.g.* urea, peroxide, borazine and L-alanine) multipole models can yield normalized mean and r.m.s. arc-length errors of better than 0.05 electrons, while for less well behaved systems typical errors of 0.05–0.15 electrons are common and in HMT the errors lie between 0.20 and 0.40 electrons.

Our final histogram (Fig. 6) provides an estimate of the r.m.s. error incurred in a Fourier summation (to $\sin\theta/\lambda = 1.0$ Å^{-1}) of the electron density or deformation density by using model phases rather than true phases. The results suggest that spherical atom models result in a error of between 0.02 to 0.06 e Å^{-3}, while multipole models will often reduce this to 0.01 e Å^{-3} or less. As before, HMT is exceptional: models *B* and *C*

incur greater r.m.s. errors than model *A* and up to 0.10 e Å^{-3}. For cyclopropane and acetamide a conventional multipole refinement (*i.e.* with fixed exponents) may still yield an r.m.s. error as large as 0.03 e Å^{-3}.

The poor performance of the more flexible models for HMT is not surprising and has been discussed in considerable detail by Terpstra *et al.* (1993), Kampermann, Ruble & Craven (1994) and El Haouzi *et al.* (1996), and is rather well understood. What our results suggest, however, is that there is a broad spectrum of behaviour of charge density analyses in non-centrosymmetric space groups. HMT is somewhat extreme and exhibits features similar to those observed in charge density analyses of several high-symmetry inorganics such as GaAs, BN in the space group $F\bar{4}3m$ (Will, Kirfel & Josten, 1986; Eichorn, Kirfel, Grochowski & Serda, 1991; El Haouzi *et al.*, 1996) and BeO in $P6_3mc$ (Downs, 1983). Cyclopropane and acetamide pose somewhat less of a problem, while for the other systems we can confidently predict that a multipole refinement will be capable of retrieval of phases to a high degree of accuracy. With reference to Table 1, we see that acetamide and cyclopropane possess substantially fewer reflections with restricted phases than all other examples (<1 and 12%, respectively). Although the importance of these restricted reflections to phasing *via* a multipole refinement will vary depending on the space group, the cell parameters and the particular classes of reflections, their number appears to be a fair indicator of success in phase retrieval.

To highlight the impact of the various statistics in Figs. 2–6, plots of the conventional Fourier deformation electron density have also been produced for urea, acetamide and HMT. The deformation densities were defined by

$$\Delta\rho(\mathbf{r}) = V^{-1}\sum_{\text{sym}}\left(|F_t|\exp(i\varphi_m) - |F_{\text{IAM}}|\exp(i\varphi_{\text{IAM}})\right)\exp(-2\pi i\mathbf{H}\cdot\mathbf{r}),$$

where the summation extends to $\sin\theta/\lambda = 1.0$ Å^{-1}. Fig. 7 displays the results for models *IAM*, *A*, *B* and *C* fitted to the dynamic data for urea, and the true (*i.e.* correctly phased) deformation density is provided for comparison; Figs. 8 and 9 provide similar maps for HMT and acetamide, respectively. (Corresponding fits to static data for these three systems yield similar results, but with more sharply defined features; we focus here on fits to dynamic data because of the direct relevance to experimental data.) The figures clearly show that the use of a spherical atom model, either *IAM* or κ

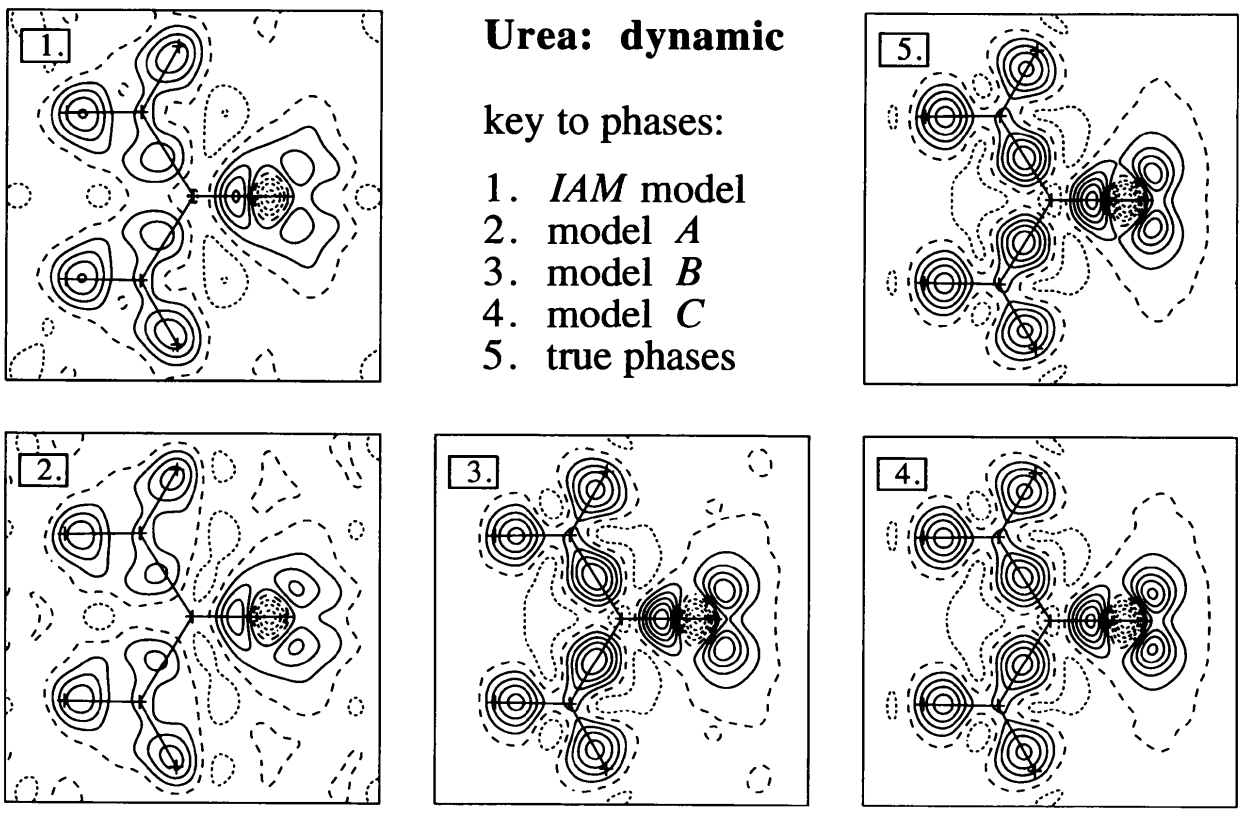

Fig. 7. A comparison of various phasing models in Fourier deformation electron densities (dynamic data) for urea in the plane of the molecule. The maps are 5 Å square, with contours at intervals of 0.10 e Å^{-3}; positive contours solid, zero dashed, negative contours dotted.

refinement, results in deformation density peak heights which are as much as 50% below the correct values. For acetamide and urea the progressive improvement $A \rightarrow B \rightarrow C$ is evident and almost quantitative agreement is observable between $\Delta\rho$ maps phased with model *C* and the true phases. For HMT that is clearly not the case and, in accordance with unconstrained analyses of experimental data (Terpstra *et al.*, 1993; Kampermann *et al.*, 1994), both multipole models result in spurious peaks inside the cage near the N atom, opposite the 'lone pair' peak, as well as peak heights in the C—H bond which are clearly too high, and corresponding deep troughs elsewhere, although model *B* is clearly worse in this regard than model *C* (in complete accord with the r.m.s. error in the electron density, Fig. 6). These maps for HMT graphically demonstrate the nature of the problem for this system: the O4 (*xyz*) octopoles on C and N are highly correlated (Terpstra *et al.*, 1993; Kampermann *et al.*, 1994) and only the use of constraints can alleviate this problem, although never remove it altogether (El Haouzi *et al.*, 1996). We explore the solution of the phase problem for HMT in more detail in the next section, but these results for HMT would seem to support the conclusion of El Haouzi *et al.* (1996) that "for non-centrosymmetric structures a 'truly experimental' electron density cannot be obtained from an X-ray diffraction experiment". However, we add the caveat that this may perhaps only be meaningful in exceptional cases such as HMT, GaAs, BN, BeO *etc.*, and that there are likely to be very many examples of molecular crystals (such as hydrogen peroxide, urea, L-alanine, borazine and even acetamide in the present work) for which the multipole model is capable of providing an extremely accurate retrieval of structure-factor phases and, hence, a very close approximation to a 'truly experimental' electron density.

5. The special case of hexamethylenetetramine

We could not conclude this study without a more detailed exploration of phasing for HMT, since it appears to represent an extreme example of phase retrieval. One important distinction between HMT and other systems studied here is that for HMT the structure-factor expression contains terms (the *xyz* octopoles) which are invariant under the symmetry operations of the space group and which result in significant changes in structure factor phases without affecting the magnitudes. Can we improve upon our

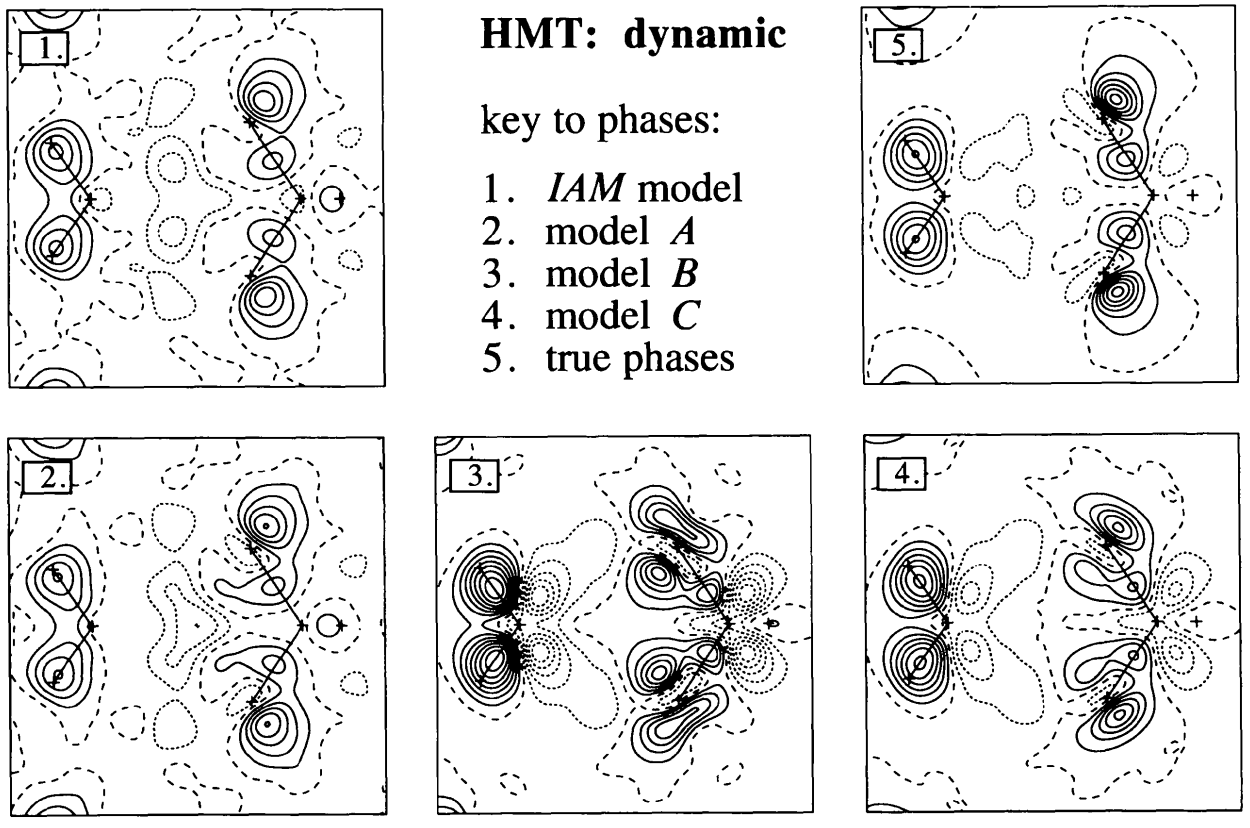

Fig. 8. A comparison of various phasing models in Fourier deformation electron densities (dynamic data) for HMT in the plane of the C—H and C—N bonds (the C—N bond is to the right). The maps are 6 Å square, with contours at intervals of 0.10 e Å^{-3}; positive contours solid, zero contours dashed, negative contours dotted.

previous results for systems like HMT? One obvious way to attempt this is to introduce constraints, for example, Terpstra *et al.* (1993) and Kampermann *et al.* (1994) set o_4 (the population of the *xyz* octopole) on C to a value typical of a tetrahedral C atom in their constrained refinements for HMT. El Haouzi *et al.* (1996) have recently discussed the automatic detection of poorly determined parameters in charge density analysis with particular reference to the phase problem for non-centrosymmetric systems and the use of 'singular value decomposition' [SVD (*e.g.* Press, Flannery, Teukolsky & Vetterling, 1986)] to alleviate the problem. In essence the approach pursued by El Haouzi *et al.* (1996) is diagonalization of the least-squares normal matrix, followed by examination of the eigenvectors corresponding to its lowest eigenvalues. This is not really SVD as we know it and as described by Press *et al.* (1986), but closer to 'eigenvalue filtering' [for a nice discussion of the differences between the two, see Watkin (1994)]. SVD and eigenvalue filtering have been incorporated into the *XD* program package (Koritsanszky, Howard, Mallinson, Su, Richter & Hansen, 1995), although we are unaware of any pubished applications to date. Our approach here is along the lines of these works and embodies eigenvalue filtering (also called characteristic-value or latent-root filtering, or principal-component analysis).

Crystallographic least-squares, especially in charge density analysis, is a non-linear process and usually involves iteratively linearizing the model structure factors and solving the normal equations to obtain a vector of parameter shifts; iteration is completed when zero shifts result. The normal equations

$$\mathbf{A}\Delta\mathbf{p} = \mathbf{b}$$

where

$$A_{ij} = \sum_{\mathbf{H}} w(\mathbf{H}) \frac{\partial|F_c(\mathbf{H})|}{\partial p_i} \frac{\partial|F_c(\mathbf{H})|}{\partial p_j}$$

and

$$b_i = \sum_{\mathbf{H}} w(\mathbf{H}) \frac{\partial|F_c(\mathbf{H})|}{\partial p_i} \left(|F_o(\mathbf{H})| - |F_c(\mathbf{H})|\right)$$

can be solved by diagonalization, or a rotation in the parameter space

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^t,$$

whereby they can be expressed in the form

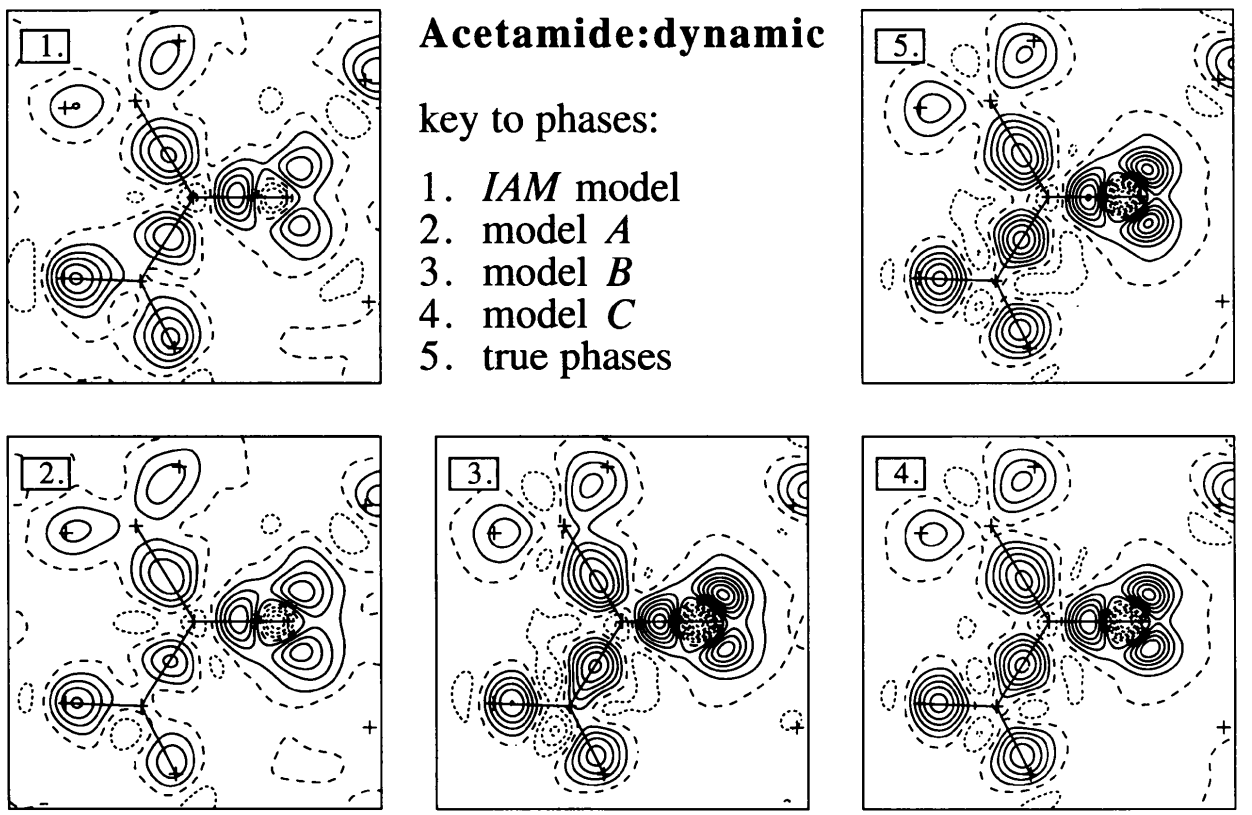

Fig. 9. A comparison of various phasing models in Fourier deformation electron densities (dynamic data) for acetamide in the plane of the peptide link. The maps are 5 Å square, with contours at intervals of 0.10 e Å^{-3}; positive contours solid, zero contours dashed, negative contours dotted.

Table 2. *Comparison of results from multipole refinements against dynamic model data for HMT: conventional unconstrained refinement versus eigenvalue filtering and eigenvalue filtering with a slack constraint on the scale factor*

Octopole populations refer to multipole functions normalized such that their absolute value integrates to 2 electrons. For the phase analysis, quantities are in the same units as in Figs. 2–6. Target values of the molecular properties from the original *ab initio* wavefunctions are $\langle r^2 \rangle = -38.83\,\mathrm{e\,\AA^2}$ and $\frac{5}{2}\langle xyz \rangle = -8.92\,\mathrm{e\,\AA^3}$.

	Conventional refinement		Eigenvalue filtering		Eigenvalue filtering with scale constraint	
Multipole model	*B*	*C*	*B*	*C*	*B*	*C*
Refinement indices						
wR (*F*) (%)	0.55	0.49	0.85	0.59	0.90	0.61
Scale factor	1.00 (2)	1.00 (5)	1.07 (5)	1.03 (2)	1.00 (1)	1.00 (1)
o_4 (C)	−0.783 (40)	−0.599 (43)	−0.166 (7)	−0.249 (0)	−0.177 (7)	−0.283 (1)
o_4 (N)	−0.331 (41)	−0.037 (57)	0.319 (15)	0.361 (1)	0.327 (15)	0.341 (1)
Molecular properties						
$\langle r^2 \rangle$ (e Å^2)	−38.0 (23)	−37.6 (35)	−57.1 (56)	−45.6 (20)	−38.2 (7)	−39.3 (4)
$\frac{5}{2}\langle xyz \rangle$ (e Å^3)	−5.3 (16)	−5.8 (21)	6.0 (23)	−4.7 (13)	−6.1 (10)	−9.1 (7)
Phase analysis						
$\langle \Delta\varphi \rangle$	−1.87	−0.80	0.69	0.61	0.59	0.44
r.m.s.($\Delta\varphi$)	3.05	1.52	2.23	1.73	2.15	1.63
$100 \times \langle \lvert F_t \rvert \Delta\varphi \rangle / F(000)$	−0.137	−0.069	0.039	0.030	0.036	0.020
$100 \times$ r.m.s.($\lvert F_t \rvert \Delta\varphi$)/$F(000)$	0.343	0.201	0.109	0.063	0.075	0.051
r.m.s.($\delta\rho$) (e Å^{-3})	0.077	0.041	0.028	0.017	0.019	0.013

$$\mathbf{D}\mathbf{U}^t\Delta\mathbf{p} = \mathbf{U}^t\mathbf{b}$$

or

$$\mathbf{D}\Delta\mathbf{p}' = \mathbf{b}'.$$

The rotation matrix **U** contains eigenvectors which relate the new set of (orthogonal) parameters, **p**′, to the old and the elements of **D**, the eigenvalues, indicate the rate of change of the residual with respect to the new parameters. $\mathbf{D}^{-1}$ is another diagonal matrix with non-zero elements just the reciprocal of those in **D** [*i.e.* $(D^{-1})_{ii} = (D_{ii})^{-1}$], hence the solution is straightforward provided no D_{ii} are zero or very close to zero (*i.e.* the original normal matrix is not 'ill-conditioned'). This last requirement is our starting point, since ill-conditioning may result from either numerical problems (or scaling) or correlations between the original parameters, the latter being our present interest. As described by Watkin (1994), if any D_{ii} are 'small', their inverse can be set to zero, hence filtering out the small eigenvalues. Shifts are then computed for the remaining parameters and the solution rotated back into the original parameter space. Thus, eigenvalue filtering in non-linear least-squares is a procedure which iteratively yields constraints on the parameter shifts, not on the parameters themselves.

El Haouzi *et al.* (1996) utilized the eigenvectors corresponding to the lowest eigenvalues of the normal matrix to deduce plausible constraints on several parameters in multipole refinements for GaAs and LiB_3O_5. We decided to try a more automated procedure and apply eigenvalue filtering automatically in refinements using our model dynamic data on HMT. For this purpose we defined a 'small' eigenvalue to be one less than 5×10^{-7} times the largest eigenvalue of the normal matrix. This is an arbitrary condition, but one which leads to filtering of only the lowest one or two eigenvalues in these refinements, and was found empirically to virtually eliminate the 'phase problem' and yield acceptable values of the properties and scale factor. The outcomes from the filtering process, especially derived physical properties of the molecular electron density and structure-factor phases, can be assessed critically in the present case because we are working with a model electron density with known properties and known structure-factor phases. Refinement details, properties and phase analysis are summarized in Table 2 for multipole models *B* and *C*, and compared with the results of the conventional refinements from the previous section.

Clearly, the use of eigenvalue filtering in this case represents a dramatic improvement in phase retrieval but, in the absence of a constraint on the overall scale factor, sometimes provides poor estimates of second and octopole moments. For this reason alone we also incorporated a 'slack constraint' (see *e.g.* Hirshfeld, 1977) on the scale factor by including $F(000)$ as an additional observation, with unit weight in this case. The phase analysis results in Table 2 are readily compared with Figs. 2–6. Eigenvalue filtering reduces mean phase angle errors to near 0.5° and below, but r.m.s. phase errors are not reduced quite as dramatically. Arc-length errors show a large reduction, however, and for model *C* with eigenvalue filtering they are now completely in accordance with the more well behaved systems such as urea, peroxide and borazine. The approximate error in the electron density is also substantially reduced, from 0.077 (conventional refinement, model *B*) to 0.013 e Å^{-3}

(eigenvalue filtering plus scale constraint, model *C*). In summary, the effect of the filtering process on the structure-factor phases is remarkable, removing all anomalous behaviour displayed by HMT in conventional refinements. Fig. 10 provides a dramatic pictorial proof of this, where Fourier deformation density maps are presented using phases from the new model *B* and *C* refinements; the maps can be compared directly with those for models *B* and *C* in Fig. 8, and with the correctly phased map in that figure. The agreement with the correctly phased map is now quite impressive, with almost quantitative agreement observed for all significant features. Importantly, eigenvalue filtering has removed all evidence of the spurious peak near the N atom on the inside of the molecular cage.

It is of some interest to examine the eigenvectors which have been filtered out of the refinements and their corresponding relative eigenvalues. Ignoring all other contributions except those involving the populations (o_4) of the *xyz* octopoles on C and N, the eigenvectors removed from the final cycle of refinement are summarized:

(i) *Eigenvalue filtering only*

Model *B*: lowest eigenvector zeroed.

Eigenvector	Relative eigenvalue
$0.814 \times o_4(\mathrm{C}) + 0.579 \times o_4(\mathrm{N})$	1.9×10^{-7}

Model *C*: lowest two eigenvectors zeroed.

Eigenvector	Relative eigenvalue
$-0.728 \times o_4(\mathrm{C}) + 0.672 \times o_4(\mathrm{N})$	2.2×10^{-7}
$0.678 \times o_4(\mathrm{C}) + 0.735 \times o_4(\mathrm{N})$	3.4×10^{-8}

(ii) *Eigenvalue filtering plus scale constraint*

Model *B*: lowest eigenvector zeroed.

Eigenvector	Relative eigenvalue
$0.808 \times o_4(\mathrm{C}) + 0.588 \times o_4(\mathrm{N})$	2.4×10^{-7}

Model *C*: lowest two eigenvectors zeroed.

Eigenvector	Relative eigenvalue
$-0.797 \times o_4(\mathrm{C}) + 0.591 \times o_4(\mathrm{N})$	3.2×10^{-7}
$0.597 \times o_4(\mathrm{C}) + 0.802 \times o_4(\mathrm{N})$	2.9×10^{-8}

As anticipated from the results for GaAs reported by El Haouzi *et al.* (1996), the lowest eigenvectors involve largely (but not entirely) the sum and difference of the two octopole populations. For model *B* only the lowest eigenvalue is removed, resulting in a shift constraint on

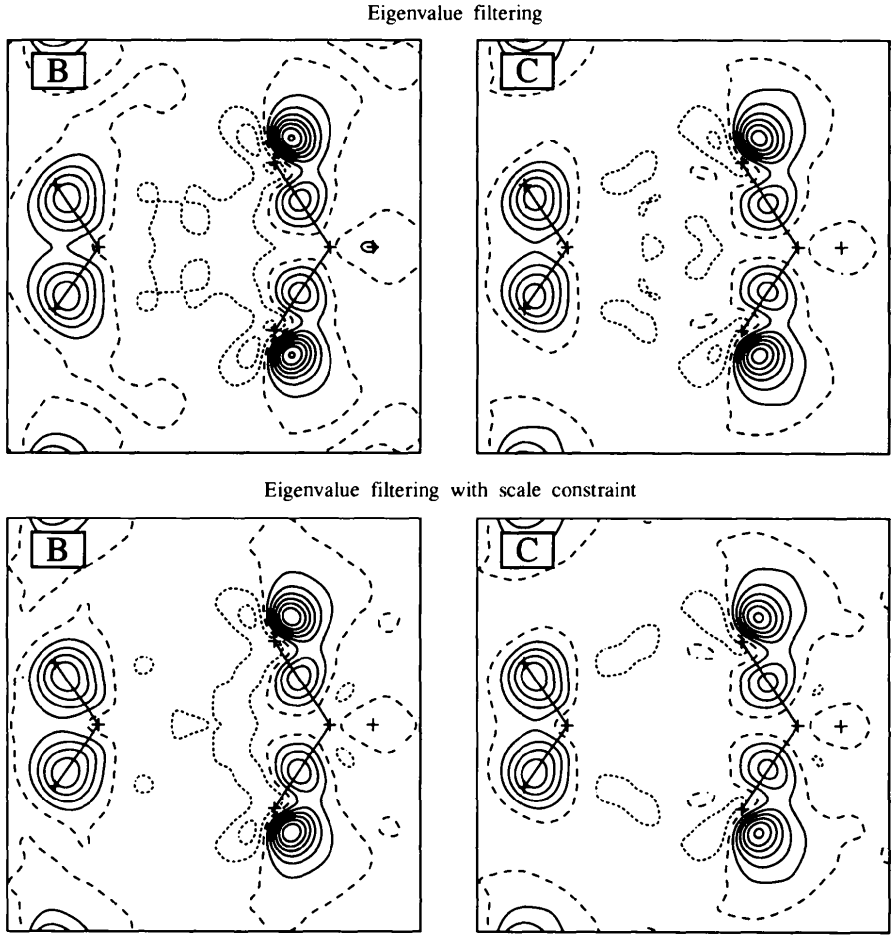

Fig. 10. The effect of eigenvalue filtering on Fourier deformation electron densities (dynamic data) for HMT. Mapping plane and contours are the same as in Fig. 8 and comparison should be made with maps 3 and 4 of Fig. 8 (which correspond to using phases from conventionally refined models *B* and *C*, respectively) and map 5 of Fig. 8 (the correctly phased deformation density).

the sum of the o_4 populations, whereas for model *C* both the sum and difference are effectively constrained, hence the e.s.d.'s of essentially zero for those refinements in Table 2. (The e.s.d.'s are not identically zero because the constraints on the shifts involve more than just the two o_4 populations). Another interesting statistic in Table 2 is the weighted *R* factor, which suggests that the cost of these constraints on the residual is substantial.

Since the eigenvalue filtering technique is applied here to non-linear least-squares, it results in constraints on shifts which are updated on each successive cycle. This implies that the final result reported in Table 2 are not unique, which is indeed the case: the results depend upon the starting parameters. For the refinements in Table 2 involving eigenvalue filtering, model *B* commenced with all populations of higher multipoles equal to zero, while model *C* commenced with the populations from model *B* (zero starting populations cannot be used with simultaneous refinement of populations and radial exponents). We believe these initial parameters are reasonable in each case and probably offer the least biased choices in the circumstances. This does raise a question about the value of the results of eigenvalue filtering quoted in Table 2 and we observe that although in each case they represent only one of many possible sets of results which may be obtained, other results obtained using *small* non-zero starting populations in model *B* are similar and yield virtually identical sets of phases.

Molecular properties derived from the multipole refined electron density are also important (Spackman & Byrom, 1996) and in Table 2 we provide estimates of the trace of the molecular second moment tensor $\langle r^2 \rangle$ and the octopole moment $\frac{5}{2}\langle xyz \rangle$ for HMT. Clearly eigenvalue filtering applied in a routine manner as carried out here can impose a constraint which leads to a significantly worse scale factor (by as much as 7% in this case) and hence compromise estimates of the molecular properties which are derived from the *rescaled* electron distribution. The estimate of the octopole moment from model *B* refinement with eigenvalue filtering has the wrong sign, but imposition of the additional slack constraint on the scale factor results in values in excellent agreement with the target *ab initio* values. In this context, it is interesting to compare the present results with values of $-3.5(5)$ (no constraint on o_4) and $-4.8(5)$ e Å^3 (constraint on o_4) from analyses of the 120 K experimental data (see Kampermann *et al.*, 1994).

A further comment emphasizing the conclusions of Terpstra *et al.* (1993) is worthwhile. These authors stated that the large least-squares correlation observed between the o_4 populations in HMT is a consequence of their small contribution to the magnitudes of the structure factors. We have confirmed this in our model studies by examining the changes in structure-factor magnitudes and phases caused by setting o_4 on both C and N to zero. We found that very few reflections display significantly large changes in magnitude (for example, compared with a typical experimental error) and those that do would be the very weakest experimentally, while phase changes of several degrees or more are commonplace. This is largely a consequence of the relatively small octopolar deformations, but it is further exacerbated by two consequences of the structure: (i) the *xyz* octopoles contribute precisely zero to the 27% of reflections with restricted phases (Table 1) and (ii) the *xyz* octopole on C affects only the imaginary part of the structure factor.

6. Conclusions

The results obtained from the model data sets used in this study may represent the limiting case with regard to what can be obtained for real data (*i.e.* the best case scenario). We emphasize that several important factors have been ignored in this study: the effect of a higher $\sin\theta/\lambda$ cut-off, the use of more flexible radial functions (or higher angular functions), the use of other weighting schemes or the influence of experimental-type errors in the least-squares process. Nevertheless, the studies which have been undertaken suggest that if sufficiently accurate experimental data can be collected, then the accurate retrieval of phases obtained in this work is likely to be replicated in experimental studies for many molecular crystals.

However, we would advise caution in all charge density studies on non-centrosymmetric crystals, especially those where high correlations are observed between multipole parameters (usually odd-order) on different atoms. In those cases constraints may in fact be necessary and we suspect that the most useful of these may be those which impose a symmetry on the local deformation functions higher than the site symmetry. Effectively this amounts to imposing our particular chemical bias on the outcome of a charge density study. This will, of course, mean that in these cases the results are no longer 'truly experimental', but nevertheless they should be of considerable use and no doubt more meaningful than those derived from an unconstrained model. An alternative would appear to be the use of eigenvalue filtering, as applied above to HMT, or some variant of that procedure. We have clear evidence that it can be highly successful in retrieving structure-factor phases with the present model data. The effects of experimental errors, both random and systematic, have not been explored, but it would seem most desirable to test its application on experimental data for HMT.

This research has been supported by the Australian Research Council. Some of the methodology contained

in this work is based on unpublished work done by MAS in collaboration with R. F. Stewart at Carnegie Mellon University.

References

Antipin, M. Y., Chernega, A. N., Lysenko, K. A., Struchkov, Y. T. & Nixon, J. F. (1995). *J. Chem. Soc. Chem. Commun.* pp. 505–506.

Belaj, F. (1992). *Acta Cryst.* B**48**, 598–604.

Blessing, R. H. & Lecomte, C. (1991). In *The Application of Charge Density Research to Chemistry and Drug Design*, edited by G. A. Jeffrey and J. F. Piniella, pp. 155–186. New York: Plenum Press.

Boese, R., Maulitz, A. H. & Stellberg, P. (1994). *Chem. Ber.* **127**, 1887–1889.

Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables* **14**, 177–478.

Coppens, P. (1974). *Acta Cryst.* B**30**, 255–261.

Coppens, P. (1982). In *Electron Distributions and the Chemical Bond*, edited by P. Coppens and M. B. Hall, pp. 61–92. New York: Plenum Press.

Coppens, P. & Hansen, N. K. (1977). *Isr. J. Chem.* **16**, 163–167.

Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* A**35**, 63–72.

Destro, R., Marsh, R. E. & Bianchi, R. (1988). *J. Phys. Chem.* **92**, 966–973.

Donohue, J. (1985). *Acta Cryst.* A**41**, 203–204.

Downs, J. W. (1983). Ph.D. Thesis. Virginia Polytechnic Institute and State University.

Eichorn, K., Kirfel, A., Grochowski, J. & Serda, P. (1991). *Acta Cryst.* B**47**, 843–848.

El Haouzi, A., Hansen, N. K., Le Hénaff, C. & Protas, J. (1996). *Acta Cryst.* A**52**, 291–301.

Feil, D. (1977). *Isr. J. Chem.* **16**, 149–153.

Fkyerat, A., Guelzim, A., Baert, F., Paulus, W., Heger, G., Zyss, J. & Périgaud, A. (1995). *Acta Cryst.* B**51**, 197–209.

Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*. Oxford University Press.

Hamzaoui, F., Baert, F. & Wojcik, G. (1996). *Acta Cryst.* B**52**, 159–164.

Hanson, J. C., Sieker, L. C. & Jensen, L. H. (1973). *Acta Cryst.* B**29**, 797–808.

Hehre, W. J., Ditchfield, R., Stewart, R. F. & Pople, J. A. (1970). *J. Chem. Phys.* **52**, 2769–2773.

Hehre, W. J., Stewart, R. F. & Pople, J. A. (1969). *J. Chem. Phys.* **51**, 2657–2664.

Hirshfeld, F. L. (1977). *Isr. J. Chem.* **16**, 226–229.

Howard, S. T., Hursthouse, M. B., Lehmann, C. W., Mallinson, P. R. & Frampton, C. S. (1992). *J. Chem. Phys.* **97**, 5616–5630.

Hümmer, K., Weckert, E. & Bondza, H. (1990). *Acta Cryst.* A**46**, 393–402.

Jeffrey, G. A., Ruble, J. R., McMullan, R. K., DeFrees, D. J., Binkley, J. S. & Pople, J. A. (1980). *Acta Cryst.* B**36**, 2292–2299.

Kampermann, S. P., Ruble, J. R. & Craven, B. M. (1994). *Acta Cryst.* B**50**, 737–741.

Koritsanszky, T., Howard, S. T., Mallinson, P. R., Su, Z., Richter, T. & Hansen, N. K. (1995). *XD. A Computer Program Package for Multipole Refinement and Analysis of Charge Densities from X-ray Diffraction Data*. Free University of Berlin, Germany.

Lecomte, C. (1991). In *The Application of Charge Density Research to Chemistry and Drug Design*, edited by G. A. Jeffrey and J. F. Piniella, pp. 121–154. New York: Plenum Press.

Lecomte, C., Ghermani, N., Pichon-Pesme, V. & Souhassou, M. (1992). *J. Mol. Struct.* **255**, 241–260.

Mighell, A. D., Himes, V. L. & Rodgers, J. R. (1983). *Acta Cryst.* A**39**, 737–740.

Mullen, D. (1980). *Acta Cryst.* B**36**, 1610–1615.

Mullen, D. (1982). *Acta Cryst.* B**38**, 2620–2625.

Mullen, D. & Scheringer, C. (1978). *Acta Cryst.* A**34**, 476–477.

Nicoud, J. F. & Twieg, R. J. (1987). In *Nonlinear Optical Properties of Organic Molecules and Crystals*, edited by D. S. Chemla and J. Zyss, pp. 227–296. Orlando: Academic Press.

Nijveldt, D. & Vos, A. (1988). *Acta Cryst.* B**44**, 281–289.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.

Radaev, S. F., Maximov, B. A., Simonov, V. I., Andreev, B. V. & D'yakov, V. A. (1992). *Acta Cryst.* B**48**, 154–160.

Savariault, J.-M. & Lehmann, M. S. (1980). *J. Am. Chem. Soc.* **102**, 1298–1303.

Shen, Q. & Colella, R. (1986). *Acta Cryst.* A**42**, 533–538.

Souhassou, M., Lecomte, C., Blessing, R. H., Aubry, A., Rohmer, M.-M., Wiest, R., Bénard, M. & Marraud, M. (1991). *Acta Cryst.* B**47**, 253–266.

Souhassou, M., Lecomte, C., Ghermani, N., Rohmer, M.-M., Wiest, R., Bénard, M. & Blessing, R. H. (1992). *J. Am. Chem. Soc.* **114**, 2371–2382.

Spackman, M. A. & Brown, A. S. (1994). *Ann. Rep. Prog. Chem. Sect. C*, **91**, 175–212.

Spackman, M. A. & Byrom, P. G. (1996). *Acta Cryst.* B**52**, 1023–1035.

Spence, J. C. H. (1993). *Acta Cryst.* A**49**, 231–260.

Srinivasan, R. (1991). *Acta Cryst.* A**47**, 452.

Stewart, R. F. (1980). In *Electron and Magnetization Densities in Molecules and Solids*, edited by P. Becker, pp. 427–431. New York: Plenum Press.

Stewart, R. F. & Spackman, M. A. (1983). *VALRAY User's Manual*. Chemistry Department, Carnegie Mellon University.

Swaminathan, S., Craven, B. M. & McMullan, R. K. (1984). *Acta Cryst.* B**40**, 300–306.

Terpstra, M., Craven, B. M. & Stewart, R. F. (1993). *Acta Cryst.* A**49**, 685–692.

Thomas, J. O. (1978). *Acta Cryst.* A**34**, 819–823.

Watkin, D. (1994). *Acta Cryst.* A**50**, 411–437.

Wiest, R., Pichon-Pesme, V., Bénard, M. & Lecomte, C. (1994). *J. Phys. Chem.* **98**, 1351–1362.

Will, G., Kirfel, A. & Josten, B. (1986). *J. Less-Common Met.* **117**, 61–71.

Zuo, J. M., Spence, J. C. H., Downs, J. & Mayer, J. (1993). *Acta Cryst.* A**49**, 422–429.

Zuo, J. M., Spence, J. C. H. & Hoier, R. (1989). *Phys. Rev. Lett.* **62**, 547–550.