Acta Cryst. (1996). B52, 1023–1035

Molecular Electric Moments and Electric Field Gradients from X-ray Diffraction Data: Model Studies

MARK A. SPACKMAN* AND PATRICK G. BYROM

Department of Chemistry, University of New England, Armidale, NSW 2351, Australia

(*Received 30 January 1996; accepted 20 June 1996*)

Abstract

Model X-ray data sets, with and without the inclusion of experimental thermal motion parameters, have been computed *via* Fourier transformation of *ab initio* molecular electron densities for 12 different molecular crystals. These datasets were then analysed with three different multipole models of varying sophistication and, from the multipole functions, molecular dipole and second moments, as well as electric field gradients (EFG's), at each nuclear site were computed and compared with results obtained from the original *ab initio* wavefunctions. The results provide valuable insight into the reliability of these properties, extracted in the same way from experimental X-ray data. Not all molecular systems display identical trends, but a general pattern is discernible. Specifically, dipole moments are typically underestimated by a small but significant amount (~10–15%), the trace of the second moment tensor is well determined but overestimated by a few per cent and electric field gradients at protons are confirmed to be well within reach of a careful charge density analysis of X-ray diffraction data.

1. Introduction

Accumulated evidence suggests that high-quality X-ray diffraction data is capable of providing estimates of the electrostatic properties of molecules in crystals, in particular, electric moments (Spackman, 1992), energies of molecular interactions (Spackman, Weber & Craven, 1988) and electric field gradients (EFG's), especially at protons (Brown & Spackman, 1994). However, few charge density analyses involve determination of these properties, partly because of the present limitations of experimental data and partly because of some skepticism that the numbers derived are meaningful. This work aims to partly redress the situation by applying a model approach to calibrate the extraction of electric moments of molecules (dipole and second moments in particular) and nuclear EFG's from sets of structure factors computed from *ab initio* molecular wavefunctions.

Our strategy is a simple one and offers the considerable advantage of being able to control a variety of variables, for example, thermal motion, data resolution and intermolecular interactions. For the present study we ignore intermolecular interactions and construct our data from a model of non-interacting molecules superimposed as they occur in the crystal. Sets of structure factors, with and without the inclusion of experimental thermal motion parameters, to a maximum $\sin\theta/\lambda = 1.0\,\text{Å}^{-1}$, are computed *via* a straightforward Fourier transformation of Gaussian orbital products for 12 different molecular crystals. These data sets are then analysed with three different multipole models of varying sophistication, with our choice of models being dictated by those currently in use for this purpose. Finally, from the multipole functions we compute molecular dipole moments, quadrupole and second moments, as well as EFG's at each nuclear site, comparing these results with those obtained from the original *ab initio* wavefunctions.

2. Computational procedure

The procedure is summarized in flowchart fashion in Fig. 1. 12 molecules were examined in the study, with crystallographic details given in Table 1. The molecular crystals were selected because they span both centrosymmetric and non-centrosymmetric space groups, a range of bonding types (from weak van der Waals – acetylene – to strong hydrogen bonding – urea), a variety of temperatures (23 K to room temperature) and thermal motion models, and most have been the subject of previous experimental charge density studies. We emphasize here that although experimental X-ray data include the effects of intermolecular interactions, our present model data sets are constructed from superimposed non-interacting molecules and hence ignore this important effect. In the future we plan to incorporate intermolecular interactions in our model data sets.

For each molecule a reference electron distribution was generated from an *ab initio* self-consistent field calculation, using nuclear coordinates obtained from X-ray or neutron positional parameters for each molecule (for details, refer to the appropriate reference in Table 1). The *GAMESS* suite of programs (Schmidt, Baldridge, Boatz, Elbert, Gordon, Jensen, Koseki,

Matsunaga, Nguyen, Su, Windus, Dupuis & Montgomery, 1993) was used with a polarized double-zeta basis set developed recently by Thakkar and co-workers (Thakkar, Koga, Saito & Hoffmeyer, 1993). This basis set (which we label DZPT) consists of the [4*s*2*p*/2*s*] contraction of a variationally improved (9*s*5*p*/4*s*) set of primitives for (Li to Ne/H), augmented with the *Pa* set of polarization functions (three Cartesian *p* functions on H, six Cartesian *d* functions on Li to Ne) described by those authors. This basis set yields results similar to the more commonly used DZP set of Dunning (1970), but with significantly lower molecular energies. Our objective was to obtain realistic reference molecular electron densities, moments and EFG's, in particular, yet keeping the basis set small to minimize the time spent in computing thousands of structure factors; the DZPT basis set represents a compromise. The *ab initio* derived dipole moments, second moments and EFG's were used as reference values in our subsequent comparisons with the results from the various multipole refinements.

For each system in Table 1 a set of unique *hkl* were generated within the limit $(\sin\theta/\lambda)_{\max} = 1.0\,\text{Å}^{-1}$ using the algorithm described by Le Page & Gabe (1979). The program *SF* (Wolff, 1995) was then used to compute two sets of structure factors from the *ab initio* charge distribution: one with (dynamic) and one without (static) thermal motion. *SF* is a rewritten version of an earlier code (*STRUFAC*) used in several previous analyses of model X-ray data generated from theory (see, for example, Spackman, 1979; Swaminathan, Craven, Spackman & Stewart, 1984; Ritchie, Cromer, Stewart, Wasserman & Ryan, 1985; Chandler & Phillips, 1986; Chandler, Figgis, Reynolds & Wolff, 1994). It incorporates an analytical Fourier transformation of Gaussian orbital products (Chandler & Spackman, 1978) and the option to include anisotropic thermal motion parameters for each atom. The treatment of thermal motion for two-centre orbital products is not unique and various possibilities have been described elsewhere (Chandler & Phillips, 1986). In the present work we chose to represent the thermal motion of the two-centre orbital product $\chi_A\chi_B$ in the form

$$T_{AB}(\mathbf{H}) = \exp[-2\pi^2 g_{AB}\mathbf{H}^T(\mathbf{U}_A + \mathbf{U}_B)\mathbf{H}],$$

where $\mathbf{U}_A$ and $\mathbf{U}_B$ are the displacement tensors of atoms *A* and *B*, respectively, $\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ and the factor g_{AB} depends on the relationship between *A* and *B*: for *A* and *B* bonded, $g_{AB} = \frac{1}{2}$ (correlated rigid-rod translation); for *A* and *B* non-bonded, $g_{AB} = \frac{1}{4}$ [uncorrelated, atoms move independently (Stewart, 1969)]. This treatment is of course an approximate one, but it has been shown that the resulting structure factors are relatively insensitive to the precise treatment of thermal motion for two-centre products (Chandler & Phillips, 1986), a consequence of the dominance of the

Table 1. *A list of the molecules used in the study, their space group and centrosymmetry, the number of reflections generated for each* $[(\sin\theta/\lambda)_{max} = 1.0\,\text{Å}^{-1}]$ *and the temperature at which the atomic positions and thermal parameters were determined [based on analysis of X-ray (X) or neutron (N) data]*

Molecule	Space group	N_o	T (K)	Reference
Formamide	$P2_1/n$ (C)	1874	90 (X)	(*a*)
Imidazole	$P2_1/c$ (C)	2903	103 (N)	(*b*)
Uracil	$P2_1/a$ (C)	3889	Room (X)	(*c*)
Acetylene	$Pa\bar{3}$ (C)	318	131 (N)	(*d*)
Benzene	*Pbca* (C)	1974	123 (N)	(*e*)
s-Triazine	$R\bar{3}c$ (C)	551	298 (N)	(*f*)
Acetamide	*R*3*c* (A)	1382	23 (N)	(*g*)
Hydrogen peroxide	$P4_12_12$ (A)	350	110 (N)	(*h*)
Urea	$P\bar{4}2_1m$ (A)	403	123 (N)	(*i*)
Borazine	$P4_32_12$ (A)	1252	115 (X)	(*j*)
Cyclopropane	$Cmc2_1$ (A)	716	94 (X)	(*k*)
Hexamethylenetetramine (HMT)	$I\bar{4}3m$ (A)	164	120 (N)	(*l*)

(*a*) Stevens (1978); (*b*) McMullan, Epstein, Ruble & Craven (1979); (*c*) Stewart & Jensen (1967); (*d*) McMullan, Kvick & Popelier (1992); (*e*) Jeffrey, Ruble, McMullan & Pople (1987); (*f*) Coppens (1967); (*g*) Jeffrey, Ruble, McMullan, DeFrees, Binkley & Pople (1980); (*h*) Savariault & Lehmann (1980); (*i*) Swaminathan, Craven & McMullan (1984); (*j*) Boese, Maulitz & Stellberg (1994); (*k*) Nijveldt & Vos (1988); (*l*) Kampermann, Ruble & Craven (1994).

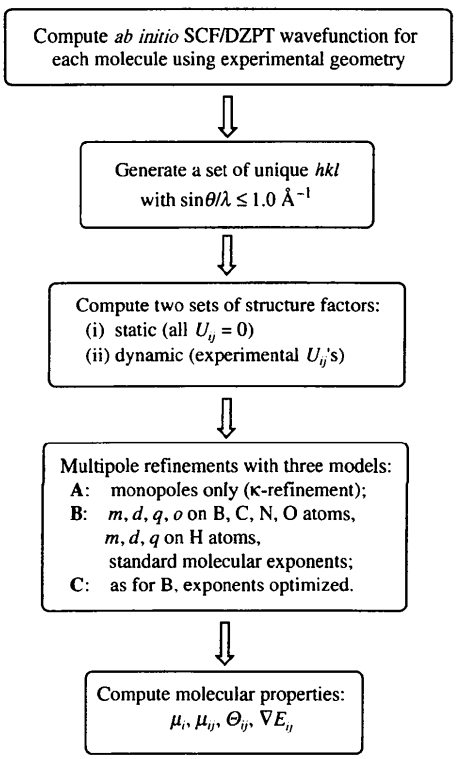

Fig. 1. A summary of the computational procedure.

contribution of one-centre products to the total X-ray scattering.

The two data sets generated for each molecular crystal were then used as the 'observed' structure factors in least-squares refinements 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). Three multipole models were refined against each data set:

(A) Monopoles only

A localized Hartree–Fock core (Clementi & Roetti, 1974; Stewart, 1980) was used for the non-H atoms with all core populations constrained to be equal. Localized Hartree–Fock 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 were 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 (Hansen & Coppens, 1978).

(B) Multipole model with fixed exponents

As for Model A plus higher multipoles on each atom (dipoles, quadrupoles and octopoles on B, C, N and 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).

(C) Multipole model with optimized exponents

As for Model B with radial parameters optimized, but constrained such that all higher-multipole radial parameters on each particular atom type are equal. For H atoms monopole exponents/scale factors were optimized independently of higher-multipole exponents.

For each of the six refinements for a particular molecular crystal (models A, B and C refined with respect to both static and dynamic data), dipole moments, second moments and EFG's at each of the nuclei were calculated by integration over the multipole functions for a single molecule removed from the crystal and compared with the reference values obtained from the original *ab initio* wavefunctions. Since the overall scale factor was not constrained in the least-squares procedure, prior to this integration the multipole population parameters were rescaled to correspond to a neutral molecule.

It is worthwhile at this point to contrast the present procedure for generating model structure factors with others employed in recent studies. Howard and co-workers have used multipole analysis of structure factors derived from *ab initio* wavefunctions to compare with experiment for (2*S*)-3-(3′,4′-dihydroxyphenyl)-alanine [L-dopa (Howard, Hursthouse, Lehmann & Poyner, 1995)] and 2-methyl-4-nitroaniline (Howard, Hursthouse, Lehmann, Mallinson & Frampton, 1992); in both cases the basis set was limited to DZ quality, the Fourier transformation was based on the method described by Stewart (1969), who provided explicit expressions only for *ss*-, *sp*- and *pp*-orbital products, and effects of thermal motion were not included. Model data sets were also computed in a similar manner for methylamine and formamide in a study of a density matrix refinement technique (Howard, Huke, Mallinson & Frampton, 1994) and for phosphoric acid in a test of the multipole procedure for phosphorus-containing systems (Moss, Souhassou, Blessing, Espinosa & Lecomte, 1995). Feil and co-workers have developed quite a different strategy, partly dictated by their choice of an STO basis set for the computation of molecular wavefunctions *via* the DVM-Xα method (Krijn, Graafsma & Feil, 1988; Velders & Feil, 1989; Bruning & Feil, 1992; De Vries, Briels & Feil, 1994; Poorthuis & Feil, 1994). Their strategy consists of numerically partitioning the theoretical electron density into atomic fragments (Hirshfeld, 1977), fitting these fragments with a finite expansion of nuclear-centred spherical harmonics with Laguerre polynomials and analytically Fourier transforming these expansions. The method has been applied to oxalic acid dihydrate (Krijn *et al.*, 1988), crystalline Si (Velders & Feil, 1989), water and its dimer (Bruning & Feil, 1992; De Vries *et al.*, 1994; Poorthuis & Feil, 1994) and, most recently, a series of nucleic acid components (Klooster, 1992). To our knowledge there has been no attempt made to compare the structure factors produced by these different procedures, although clearly that would be of some interest.

Very few of the studies cited above paid attention to the extraction of molecular electric moments from the simulated X-ray data; none addressed the extraction of EFG's. Both studies which determined dipole moments (Howard *et al.*, 1992; Howard *et al.*, 1995) ignored thermal motion completely (*i.e.* the simulated data was for a static electron distribution), which, as we shall demonstrate below, leads to a significant bias in the outcomes. Finally, only one previous study of this kind

has addressed quadrupole moments, that by Feil & Moss (1983) on pyrazine and in that instance both static and dynamic structure factors were employed, although the latter were based on a simple overall isotropic thermal motion model. There has, however, been a single direct space study which explored the effects of molecular proximity (*i.e.* overlap of electron distributions) on dipole moments of formamide derived by partitioning the electron density of the crystal or procrystal (Moss & Coppens, 1980). That study concluded that the reasonable agreement obtained between gas phase and solid-state dipole moments appears to result from an approximate cancellation of two effects: overlap in the crystal and intermolecular interactions. Although the objectives of that study overlap with those of the present work, the conclusions reached therein were based on a direct space partitioning of electron densities and as such cannot be compared directly with our own results.

3. Results and discussion

Our results will be examined under several headings. The first section will discuss the quality of the fit obtained to the simulated diffraction data by examining two residuals. Then the values of the radial function exponents derived from the refinement will be investigated. Finally, a comparison will be made between the electrical properties derived from the multipole refinements and the reference *ab initio* properties.

3.1. *Residuals and goodness-of-fit*

Tables 2 and 3 give final residuals, R_F, and goodness-of-fit, S, indices for individual multipole refinements for each molecular system, as well as mean values for centrosymmetric and non-centrosymmetric structures. Although our sample of 12 different structures is a small one, averages listed in Table 2 are meaningful and support the expectation that residuals are generally lower for non-centrosymmetric structures, rationalized by the additional degree of freedom provided by variable phases, although exceptions are evident in the table. Goodness-of-fit indices, on the other hand, display little difference between non-centrosymmetric and centrosymmetric structures. Differences are evident, however, in the quality-of-fit obtained to static *versus* dynamic structure factors. The κ-refinement model (model A) typically yields a lower residual for static data, yet a higher goodness-of-fit; higher-multipole models (B and C) almost invariably fit dynamic data better than the static case. An explanation for this behaviour lies partly in our use of unit weights in the least-squares. Incorporation of thermal motion effectively down-weights the higher-angle data (*i.e.* core regions and sharp deformations), allowing the pseudo-atom model to minimize residuals for the lower angle data (*i.e.* valence regions) where structure factors will be relatively large, at the expense of the higher-angle data (this is discussed in more detail below). Furthermore, the optimization of radial function exponents, which is the only difference between models B and C, introduces only a few more parameters into the refinement process, but

Table 2. *Percentage residuals,* $\%R_F = 100 \times \Sigma\big||F_o| - |F_c|\big| / \Sigma|F_o|$, *after multipole refinements with models A, B and C against static and dynamic data for each of the 12 molecular systems*

	Static			Dynamic		
Molecule	A	B	C	A	B	C
Centrosymmetric space groups						
Formamide	2.63	0.84	0.53	2.99	0.84	0.49
Imidazole	2.38	0.83	0.63	3.01	0.87	0.63
Uracil	2.67	0.89	0.58	3.51	0.76	0.41
Acetylene	2.60	0.94	0.54	4.07	0.56	0.27
Benzene	2.83	1.34	1.17	3.55	1.22	0.92
s-Triazine	2.84	0.85	0.51	2.98	0.85	0.51
Mean	2.66	0.95	0.66	3.35	0.85	0.54
Non-centrosymmetric space groups						
Acetamide	1.92	0.58	0.37	2.02	0.59	0.35
Hydrogen peroxide	3.42	0.67	0.55	3.59	0.67	0.55
Urea	1.88	0.60	0.35	2.02	0.62	0.31
Borazine	1.96	0.57	0.36	2.71	0.56	0.30
Cyclopropane	1.73	0.51	0.38	2.31	0.47	0.31
HMT	2.45	0.59	0.55	2.91	0.58	0.53
Mean	2.23	0.59	0.43	2.59	0.58	0.39

Table 3. *Goodness-of-fit,* $S = [\Sigma(|F_o| - |F_c|)^2 / (N_o - N_p)]^{1/2}$, *after multipole refinements with models A, B and C against static and dynamic data for each of the 12 molecular systems*

The number of parameters, N_p, is given for each of models A, B and C, respectively, and is the same for fits to both static and dynamic data for the same system (N_o is given in Table 1).

		Static			Dynamic		
Molecule	N_p	A	B	C	A	B	C
Centrosymmetric space groups							
Formamide	11, 80, 84	0.195	0.053	0.033	0.130	0.029	0.017
Imidazole	13, 120, 123	0.238	0.062	0.047	0.170	0.039	0.028
Uracil	17, 169, 173	0.307	0.087	0.055	0.066	0.012	0.007
Acetylene	5, 12, 14	0.187	0.044	0.024	0.077	0.007	0.004
Benzene	9, 78, 80	0.290	0.103	0.091	0.182	0.043	0.031
s-Triazine	7, 25, 28	0.708	0.153	0.092	0.658	0.136	0.081
Mean		0.321	0.084	0.057	0.214	0.044	0.028
Non-centrosymmetric space groups							
Acetamide	14, 114, 118	0.659	0.166	0.104	0.543	0.129	0.076
Hydrogen peroxide	5, 28, 30	0.237	0.045	0.037	0.185	0.032	0.027
Urea	10, 39, 43	0.136	0.035	0.022	0.101	0.022	0.012
Borazine	12, 80, 83	0.232	0.049	0.031	0.147	0.024	0.012
Cyclopropane	9, 59, 61	0.211	0.042	0.032	0.119	0.018	0.012
HMT	7, 21, 24	0.385	0.082	0.077	0.273	0.051	0.046
Mean		0.310	0.070	0.051	0.228	0.046	0.030

results in a highly significant improvement of the fit to the simulated data.

Typical residual maps are provided in Fig. 2 for formamide. These maps reinforce the conclusion that fits to static data are systematically worse than fits to dynamic data, but also provide a pictorial impression of precisely what R_F values of less than 0.01, and S less than 0.1 mean in the present context. To summarize for formamide:

B:static - $R_F = 0.0084$, $S = 0.053$ correspond to significant residual features (typically between 0.10 and 0.30 e Å^{-3}), especially in the vicinity of the C=O and C—N bonds.

C:static - $R_F = 0.0053$, $S = 0.033$ correspond to much smaller residual features, typically less than 0.15 e Å^{-3}.

B:dynamic - $R_F = 0.0084$, $S = 0.029$ correspond to features of similar magnitude to C:static and typically less than 0.15 e Å^{-3}.

C:dynamic - $R_F = 0.0049$, $S = 0.017$ correspond to a virtually flat residual map, almost no features greater than 0.05 e Å^{-3}.

It is worthwhile comparing the features in these maps with the residual map reported by Howard *et al.* (1995) for L-dopa. In that map, for which the multipole refinement corresponds very closely to C:static, there are features in the vicinity of bonds of magnitude 0.05–0.10 e Å^{-3}, as well as negative density (about 0.30 e Å^{-3}) around each of the heavy atom nuclei. Howard *et al.* attributed these systematic negative features to the difference between Hartree–Fock core functions and the corresponding description with a DZ Gaussian basis set (used to construct the model data). If this was a valid explanation the features would also be present in Fig. 2. Although there are small differences between the monopole functions used in the two studies [*e.g.* the present work employed localized core and valence functions, while Howard *et al.* used core and valence scattering factors from *International Tables for X-ray Crystallography* (Cromer, 1974), which should also reproduce the Hartree–Fock atomic scattering factors] and in their treatment (*e.g.* variable cores in the present work, while Howard *et al.* fixed core populations at 2.0), we can find no evidence that these differences are the cause of the negative features. We suspect that they arise from the use of an inappropriate valence scattering factor by Howard *et al.*, namely just the $\langle j_0 \rangle_{2s}$ scattering factor, rather than an appropriate linear combination of $\langle j_0 \rangle_{2s}$ and $\langle j_0 \rangle_{2p}$.

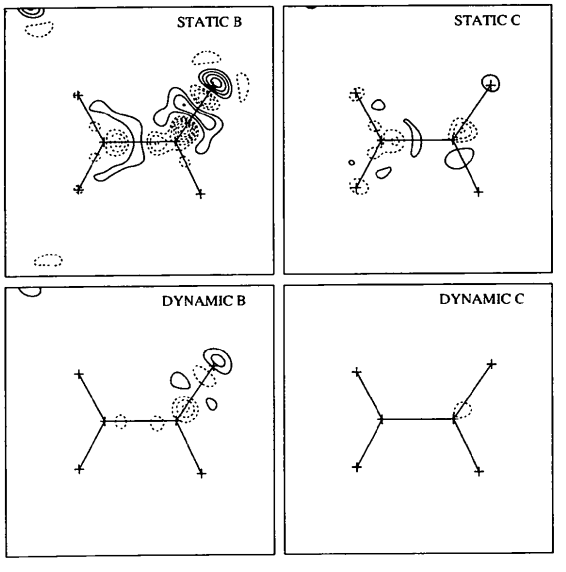

Fig. 2. Residual maps after fits to static and dynamic model data for formamide with multipole models B and C. Each map is 5 Å square in the molecular plane with the N→C vector directed left to right and the C=O bond top-right. The contour interval is 0.05 e Å^{-3} with negative contours dashed and zero contour suppressed.

Table 4. *Mean values of the optimized exponents (in a.u.) for single exponential radial functions for each atom, compared with the standard molecular (Hehre et al., 1969, 1970) and single-zeta atomic values (Clementi & Roetti, 1974)*

The mean values for both the static and dynamic refinements are shown, with standard deviations of the distribution (except for boron, which is obtained from only one molecule). For H, results are also given for monopole functions.

Atom		Static	Dynamic	Standard molecular	Atomic single-zeta
H	Monopole	A: 2.42 (20)	A: 2.80 (55)		
		B: 2.25 (7)	B: 2.32 (8)		
		C: 2.33 (11)	C: 2.42 (11)		
	Higher multipole	2.66 (23)	2.83 (31)	2.48	2.00
B		2.44	2.47	3.00	2.50
C		2.81 (16)	2.82 (13)	3.44	3.18
N		3.89 (25)	3.71 (36)	3.90	3.84
O		5.04 (13)	4.88 (27)	4.50	4.47

Although not presented in detailed tabular form, the estimated scale factors after multipole refinement (*i.e.* the multipole estimate of the total number of electrons divided by the correct number for a neutral molecule) do provide a useful indication of the quality of the fit. For model A scale factors are typically within 4% of unity and better for fits to dynamic data (static: mean Δ of 0.042, r.m.s. Δ of 0.048; dynamic: mean Δ of 0.017, r.m.s. Δ of 0.027). Much improved results are obtained for models B and C, where the difference from unity is typically less than 1% in both cases (model B:static: mean Δ of 0.011, r.m.s. Δ of 0.012; dynamic: mean Δ of 0.004, r.m.s. Δ of 0.006; model C:static: mean Δ of 0.001, r.m.s. Δ of 0.006; dynamic: mean Δ of 0.000, r.m.s. Δ of 0.006).

3.2. *Radial exponents and kappas*

Mean optimized values of the higher multipole exponents for all atoms are given in Table 4, along

with mean values of the monopole exponent for H atoms. Comparison with SM (Hehre *et al.*, 1969, 1970) and single-zeta atomic values (Clementi & Roetti, 1974) is also made in the table, although this requires some caution, as the others have been obtained by energy minimization, while our procedure involves a fit to the electron density.

Although not an objective of this study, the results in Table 4 convincingly illustrate that mutipole refinement of simulated X-ray data is a feasible route to the determination of realistic radial function exponents for use in studies of experimental data (for example, see the recent study by Moss *et al.*, 1995). For models B and C it is clear that optimum exponents are less than SM values, which suffice for N, and for O the optimum exponent is significantly larger than the SM value. In the case of H, higher multipole exponents are somewhat greater than the SM value of 2.48 and with a wide range in values observed for the different molecules; monopole exponents are close to the SM value, except for refinements with dynamic data using monopoles only. The e.s.d. for the distribution, given in parentheses in Table 4, suggests that optimum exponents for C atoms vary little, irrespective of the type of C atom [our sample of molecules contains carbonyl, C(aromatic), Csp^3, Csp^2 and Csp], while exponents for N atoms span a wide range [our sample contains N(aromatic), Nsp^2 and Nsp^3].

3.3. *Dipole moments*

The dipole moment is a vector quantity and requires three components for complete specification; for the six dipolar molecules in our study, this would require 36 sets of three components. For our present purpose, not all of this information is essential and we present our results for this property (and others below) in an abbreviated fashion, in this case concentrating only on the magnitude of the dipole moment. Molecular dipole moment magnitudes (with least-squares-derived error bars) for the six dipolar molecules are plotted in Fig. 3, along with the benchmark *ab initio* values and differences between the two. In addition, Table 5 summarizes deviations from target values (both mean and r.m.s. deviations). Important conclusions emerging from Fig. 3 and Table 5 are:

(i) Dipole moments are systematically underestimated by a small amount. This is clearly seen in the uniformly positive sign of the mean deviation and

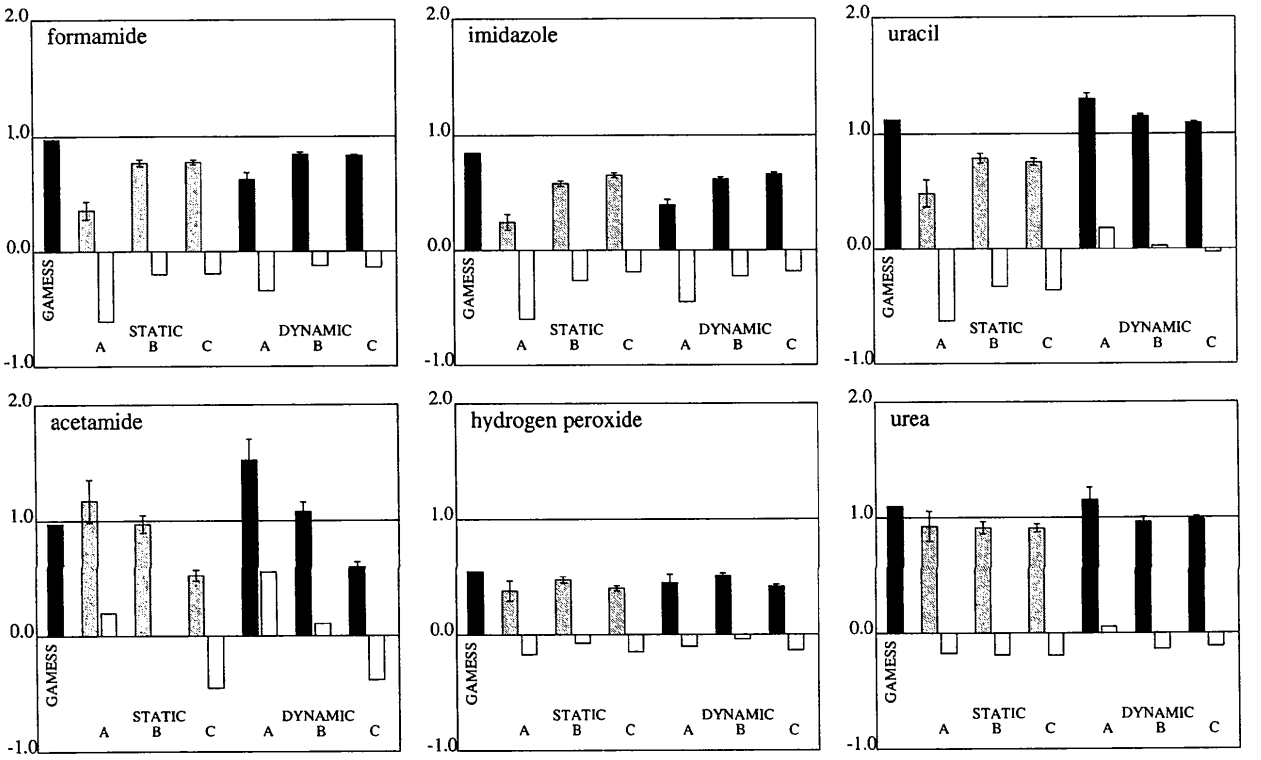

Fig. 3. Dipole moment magnitudes (e Å) with associated least-squares e.s.d.'s for each of the three multipole models (A, B and C) refined against both static (light grey bars) and dynamic (dark grey bars) data. The benchmark value from *GAMESS* is indicated by the black bar on the left and differences between each multipole result and the benchmark value are shown as a white bar. Only results for dipolar molecules are displayed.

Table 5. *Summary of mean deviation and r.m.s. deviation of molecular moments from benchmark values; units for the properties are the same as those in Figs. 3–6*

The deviation Δ is defined by $\Delta = P_{\text{ab initio}} - P_{\text{multipole}}$ for property P and σ is the least-squares derived e.s.d. in $P_{\text{multipole}}$.*

	Static			Dynamic		
	A	B	C	A	B	C
Dipole moment magnitude						
Mean Δ	0.331	0.176	0.256	0.018	0.065	0.164
R.m.s. Δ	0.453	0.207	0.279	0.335	0.129	0.196
Mean $\%\Delta$	36.1	18.6	27.6	4.8	7.6	18.8
R.m.s. $\%\Delta$	47.9	21.3	29.2	36.4	14.2	22.1
Mean Δ/σ	4.0	5.4	8.8	1.4	3.4	7.6
R.m.s. Δ/σ	5.4	6.5	9.1	4.7	6.0	8.4
$\langle r^2 \rangle$						
Mean Δ	2.820	0.850	0.268	0.885	0.310	0.232
R.m.s. Δ	3.437	1.098	0.725	1.896	0.702	0.716
Mean $\%\Delta$	−17.0	−5.5	−1.9	−4.9	−2.0	−1.9
R.m.s. $\%\Delta$	19.6	6.7	4.5	9.8	3.5	4.4
Mean Δ/σ	6.1	5.0	1.6	1.0	1.4	2.9
R.m.s. Δ/σ	6.9	6.1	4.1	6.2	6.0	7.1
Out-of-plane second moment						
Mean Δ	0.598	0.267	0.017	0.104	0.100	0.049
R.m.s. Δ	0.740	0.337	0.177	0.442	0.195	0.224
Mean $\%\Delta$	−11.0	−4.8	−0.6	−3.2	−1.9	−1.1
R.m.s. $\%\Delta$	16.5	6.9	3.8	8.3	3.5	4.5
In-plane second moment						
Mean Δ	1.349	0.353	0.138	0.534	0.135	0.094
R.m.s. Δ	1.596	0.463	0.339	0.902	0.311	0.299
Mean $\%\Delta$	−23.4	−6.6	−2.8	−9.8	−2.7	−2.0
R.m.s. $\%\Delta$	26.6	8.0	5.4	13.8	4.5	4.8

* The averages refer to the following sets of molecules:
dipole moment magnitude: formamide, imidazole, uracil, acetamide, hydrogen peroxide, urea;
$\langle r^2 \rangle$: formamide, imidazole, uracil, acetylene, benzene, *s*-triazine, acetamide, hydrogen peroxide, urea, borazine, cyclopropane;
Out-of-plane and in-plane second moments: formamide, imidazole, uracil, benzene, *s*-triazine, acetamide, urea, borazine, cyclopropane.

this underestimate is typically between three and nine times the corresponding least-squares error (or 10–25% of the target value) and hence highly significant.

(ii) The κ-refinement model (A) is usually inferior compared with higher multipole models and of these model B yields better overall agreement with reference results. Table 5 illustrates this most clearly in the figures for r.m.s. deviations.

(iii) Better estimates are generally obtained from dynamic data than from static data, as evidenced by lower deviations for all models fitted to dynamic data.

(iv) For acetamide, optimization of radial exponents (model C) dramatically worsens the estimate, no doubt a consequence of the non-centrosymmetric space group, although similar behaviour is not observed for urea and hydrogen peroxide.

We will defer discussion of the implications of these observations until after the presentation of results for second moments and EFG's.

3.4. *Second moments*

As the second moment tensor generally requires six components for complete specification, we focus this examination of results on its trace, $\langle r^2 \rangle$ [which is related to the bulk diamagnetic susceptibility, see, for example, Hinchliffe (1987), and the average Coulomb potential (Φ_0), see, for example, O'Keeffe & Spence (1994)], and on the in-plane and out-of-plane components for the nine molecules possessing a readily identifiable plane of symmetry. Fig. 4 plots the trace of the second moment tensor, with least-squares derived error bars and the differences from the benchmark *GAMESS* value, for all 12 molecules in our study; Table 5 summarizes deviations numerically. There is a general trend for most models to overestimate this quantity by a small amount, especially the monopole models (note that the quantity $\langle r^2 \rangle$ possesses a negative sign, while Fig. 4 plots only the magnitude). This large overestimate for model A is a direct consequence of its similarity to the promolecule, a model of overlapping spherical atoms, for which the mean and r.m.s. deviations comparable to the values reported in Table 5 are 3.25 and 3.78 e Å^2, respectively, and mean and r.m.s. per cent deviations are −22.0 and 24.2%, respectively, all considerably greater than the model A results reported in Table 5. This observation, that chemical bonding results in a systematic reduction in the magnitude of $\langle r^2 \rangle$ compared with isolated atoms, is in accord with the conclusions of O'Keeffe & Spence (1994).

As for the dipole moment, better agreement with target *ab initio* values is generally found for fits to the dynamic data and for models B and C refined against dynamic data there is virtually exact agreeement with *ab initio* values within the least-squares errors (especially notable for urea, formamide and acetylene), with the exception of borazine, uracil and cyclopropane. We conclude that the expectation value of r^2 is a well determined quantity from multipole refinements of this kind, although there is a systematic tendency for the multipole result to overestimate the true value. {We note that O'Keeffe & Spence (1994) reported that for Si the pseudo-atom result [−10.16 (3) e Å^2] is completely at odds with a recent electron holography measurement which implies $\langle r^2 \rangle = -6.15\,(5)$ e Å^2 (Gajdardziska-Josifovska, McCartney, De Ruijter, Smith, Weiss & Zuo, 1993) and actually greater in magnitude than the atomic value of −9.03 e Å^2. Compared with this large discrepancy, the present results demonstrate that the pseudo-atom model is capable of retrieving this quantity from X-ray diffraction data for molecular crystals.}

In order to explore the ability of the multipole models to retrieve the anisotropy of the molecular second moment tensor (upon which the quadrupole moment tensor depends in a very subtle manner) we have diagonalized the second moment tensor for nine

molecules possessing an exact or approximate mirror plane of symmetry; Fig. 5 presents the resulting out-of-plane principal components, Fig. 6 presents the mean of the two in-plane principal components. Estimated approximate errors in the derived properties are also presented in these figures and these have been based upon a Monte Carlo (Boswell, Gore, Patil & Taillie, 1993) variation of multipole parameters within their least-squares derived e.s.d.'s, since the derivation of e.s.d.'s in principal components of the tensor from e.s.d.'s in individual components before diagonalization is a non-trivial task. Careful inspection of these two figures, and the summary in Table 5, shows that the differences from benchmark results are significantly smaller for the out-of-plane components, especially noticeable for fits to the static structure factors and this is further supported by the mean and r.m.s. deviations listed in Table 5. We attribute this to the different origin of contributions to the two terms. The only contributions to the out-of-plane molecular moments are atomic second moments derived from the monopole and quadrupole functions; in contrast, the in-plane molecular moments depend upon these as well as net atomic charges (multiplied by the square of their

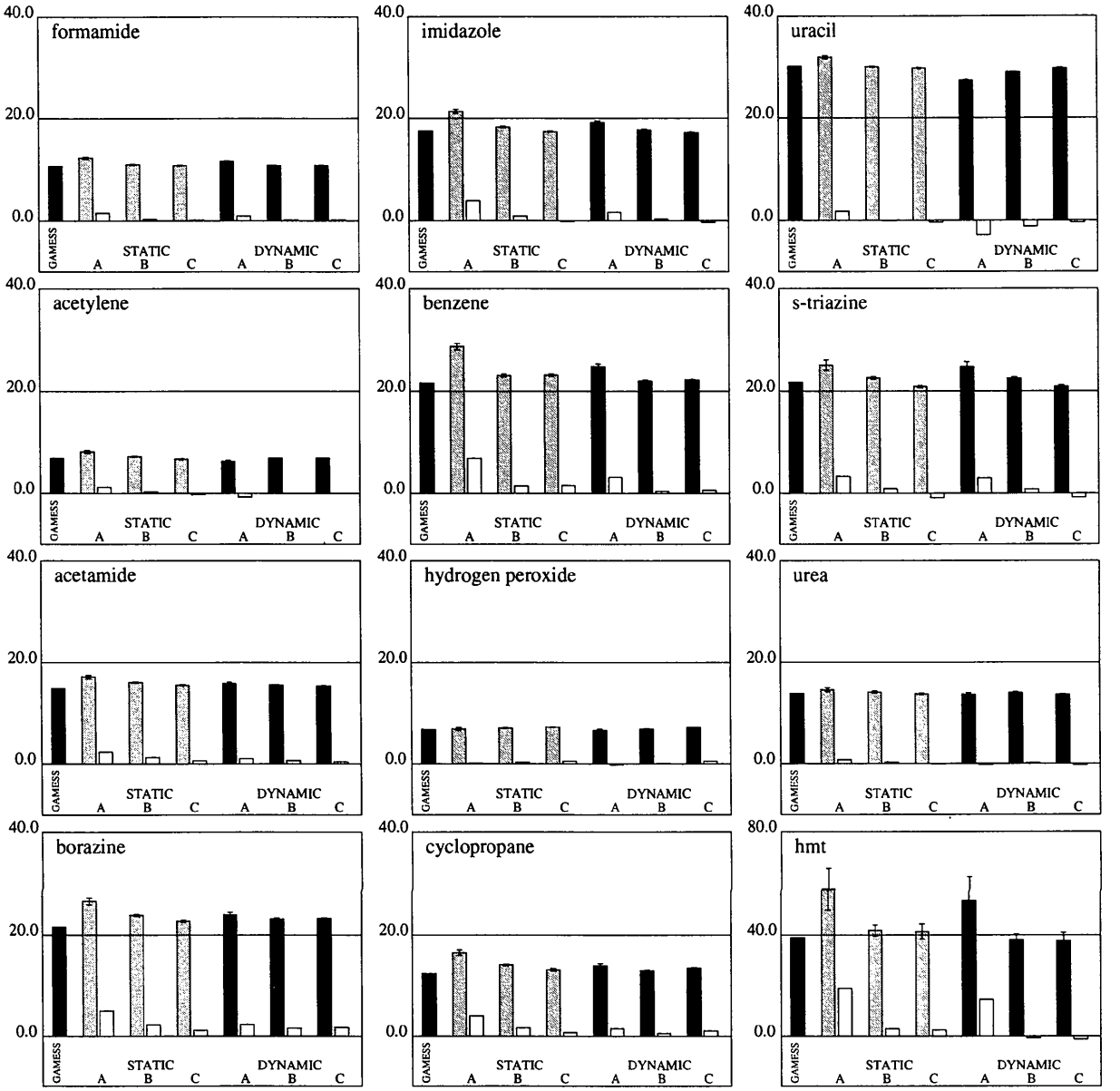

Fig. 4. Negative of the trace of the second moment tensor, $-\langle r^2 \rangle$ (e Å^2), with associated least-squares e.s.d.'s; refer to Fig. 3 for key to shading. Results for all 12 molecules are displayed; note that the vertical scale for HMT is twice that of the other histograms.

distance from the centre-of-mass origin) and net atomic dipole moments (multiplied by their distance from the origin). In a sense the out-of-plane moments are less susceptible to variations in the parameters in the multipole model and this is entirely in accord with results of analyses of experimental data for urea, benzene and *p*-dicyanotetrafluorobenzene (Spackman, 1992), where agreement between experiment and theory was also observed to be much better for out-of-plane than for in-plane moments.

3.5. *Electric field gradients*

The EFG is also a symmetric tensor, but since our interest here is in general trends, we focus on just the principal component with the largest magnitude (conventionally labelled ∇E_{zz}) and its comparison with the same quantity obtained directly from the *ab initio* wavefunction. For each heavy atom (B, C, N and O) there is a range of ∇E_{zz} values due to different chemical environments and we have computed the mean and standard deviation for these distributions, the results being plotted in Fig. 7. For H atoms we computed separate means over the range for H atoms bonded to different heavy atoms; thus, Fig. 8 displays mean values and e.s.d.'s for H(B), H(C), H(N) and a single result for H(O) from hydrogen peroxide.

The monopole only model (A) is clearly inadequate, except for H. This is neither surprising nor new, as a set of monopoles cannot be expected to reproduce adequately the sharp quadrupolar deformation near a heavy-atom nucleus, but for H atoms it is clear that the conclusion obtained by Brown & Spackman (1994) is quite general: the EFG at H is dominated by the monopole function on the nearest-neighbour atom. From Fig. 8, multipole models B and C represent quantitative retrieval of EFG tensors for H atoms, even where the structure factors include (known) thermal motion. This does not necessarily mean that all components of the tensor are retrieved equally well, or that sensitive parameters such as the anisotropy η are well determined; these details will be the subject of a future in-depth study.

For multipole models applied to N and O atoms, ∇E_{zz} is underestimated by a large and significant amount (typically 30% or more), while ∇E_{zz} for B and C can often be predicted accurately, particularly if the exponents of the radial functions are optimized. This

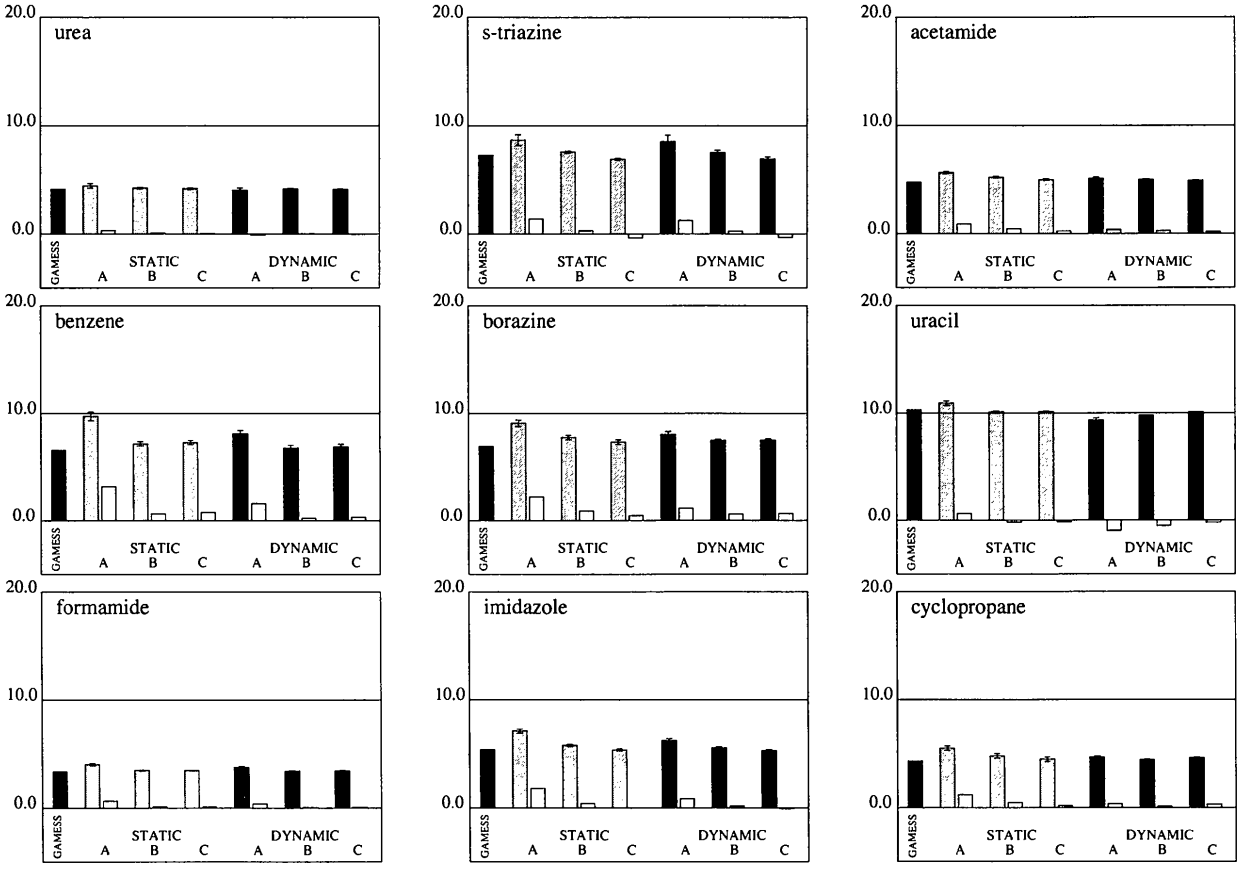

Fig. 5. Negative of the out-of-plane component of the second moment tensor (e Å^2), with estimated e.s.d.'s; refer to Fig. 3 for key to shading. Results for nine approximately planar molecules are displayed.

behaviour is almost certainly a reflection of the sharpness of the important local quadrupole deformation (or polarization) at the nucleus of interest: for B and C this deformation is relatively diffuse and most of the necessary information is contained in the structure factors, while for N and O the quadrupole deformation is increasingly sharp and its Fourier transform truncated within the present sphere of data. We would expect the situation for F and second-row atoms to be substantially worse, although we hesitate to draw any firm conclusions before more flexible radial functions are tested.

4. Fits to static *versus* dynamic data

We have remarked above on several occasions that estimates of dipole or second moments derived from multipole refinement are typically better for fits to dynamic data than to static data. This has also been observed to be the case in studies by Klooster (1992), involving the extraction of the electrostatic potential from model data sets akin to those in the present work, and has been rationalized in terms of the down-weighting of higher-angle data upon inclusion of thermal motion. This rationale can be placed on a firmer footing by comparing several effective weighting functions in the following residuals, in each case assuming static model F_o:

(i) Conventional refinements on $|F|$ (*i.e.* fitting the electron distribution)

$$\varepsilon_{\rm i} = \Sigma(|F_o| - |F_c|)^2.$$

(ii) Fitting the electrostatic potential

$$\varepsilon_{\rm ii} = \Sigma[|F_o|/(\sin\theta/\lambda)^2 - |F_c|/(\sin\theta/\lambda)^2]^2.$$

(iii) As for (i) but applying overall isotropic thermal motion to F_o and F_c

$$\varepsilon_{\rm iii} = \Sigma[|F_o|e^{-B(\sin\theta/\lambda)^2} - |F_c|e^{-B(\sin\theta/\lambda)^2}]^2.$$

(iv) As for (i) but using a data cutoff

$$\varepsilon_{\rm iv} = \Sigma w(|F_o| - |F_c|)^2,$$

where $w = 1.0$ for $\sin\theta/\lambda < (\sin\theta/\lambda)_{\rm max}$, zero otherwise.

In Fig. 9 we compare the weighting functions in cases (ii) (arbitrarily scaled by 0.15 to bring it in line with the others), (iii) (using an overall isotropic $B = 2.0\,\text{Å}^2$, appropriate to data collected at ~100 K) and (iv) [with $(\sin\theta/\lambda)_{\rm max} = 0.6\,\text{Å}^{-1}$]. Evidently schemes (iii) and (iv) are similar and (ii) differs greatly at lower $\sin\theta/\lambda$.

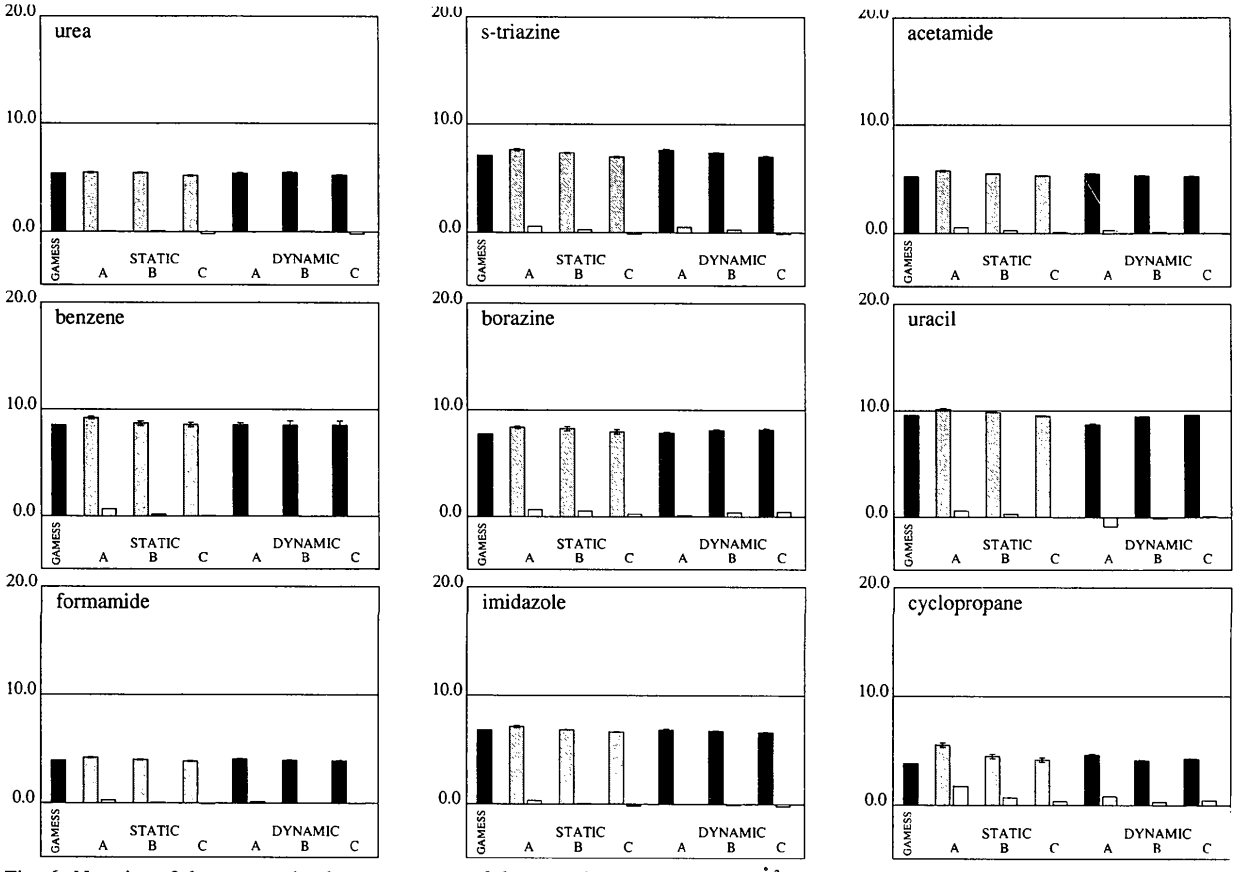

Fig. 6. Negative of the average in-plane component of the second moment tensor (e Å^2), with estimated e.s.d.'s; refer to Fig. 3 for key to shading. Results for nine approximately planar molecules are displayed.

Results of refinements employing these weighting schemes (Table 6) on the dipole moment for formamide strongly suggest that strategies (iii) and (iv) yield similar results, both strikingly close to the results derived from refinements against the dynamic data (*i.e.* that constructed using individual anisotropic thermal parameters). The point here is that multipole refinements using static model data yield electric moments (and by inference electrostatic potentials) which differ systematically from those of a fit to corresponding dynamic data and this is largely a consequence of the effective down-weighting of higher-angle data. For this reason, critical comparisons between theory and experiment based upon sets of static structure factors, as employed recently by Howard *et al.* (1992, 1995), should be treated with caution and we would strongly recommend that some form of realistic thermal motion be included in the model datasets in the future. It is perhaps fortunate, in the present context, that nature provides us with X-ray diffraction data incorporating thermal motion effects, thereby ensuring that conventional least-squares multipole refinements are biased in favour of the successful extraction of properties such as the electrostatic potential and dipole and second moments.

5. Final remarks

The present model study suffers from many obvious deficiencies: only a small number of systems were studied; only a single basis set and level of theory was used to construct model data sets; only three simple

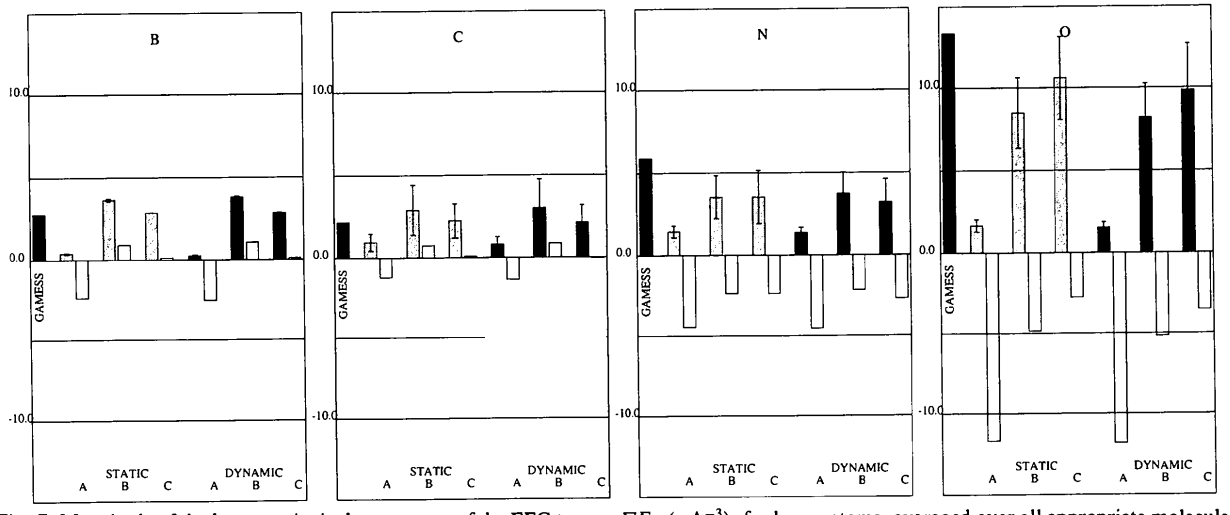

Fig. 7. Magnitude of the largest principal component of the EFG tensor, ∇E_{zz} (e A^{-3}), for heavy atoms, averaged over all appropriate molecules in the study; the error bars are standard deviations associated with the mean.

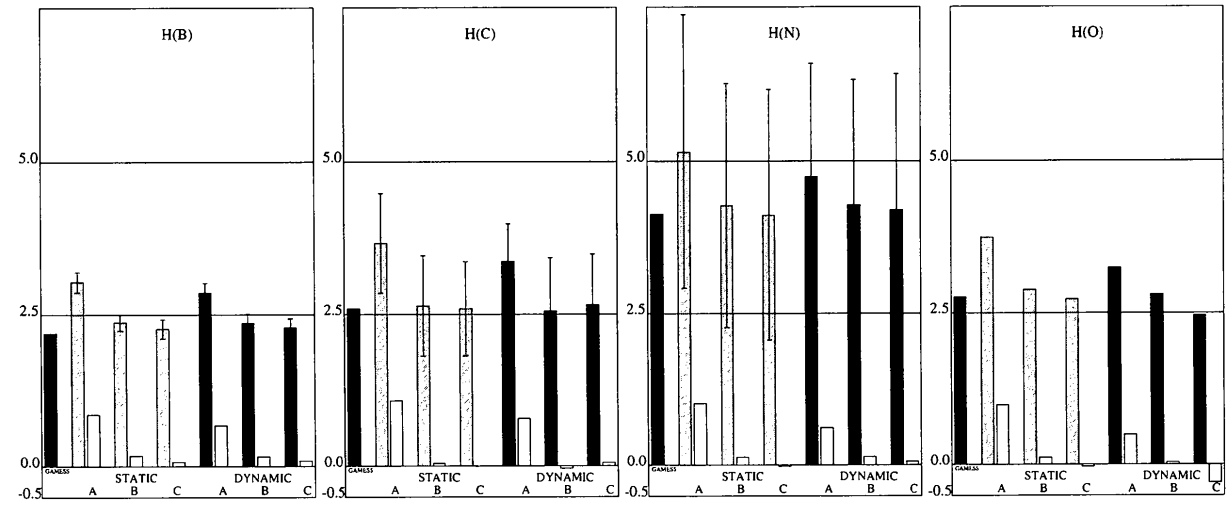

Fig. 8. Magnitude of the largest principal component of the EFG tensor, ∇E_{zz} (e $Å^{-3}$), for H atoms bonded to different types of heavy atoms, averaged over all appropriate molecules in the study; the error bars are standard deviations associated with the mean.

Table 6. *Dipole moment magnitudes (e Å) for formamide obtained by multipole refinement with various weighting schemes (see text for details)*

	Type of refinement	Multipole model A	B	C
(i)	Static, fit to electron density	0.360	0.771	0.777
(ii)	Static, fit to electrostatic potential	0.781	0.852	0.879
(iii)	Dynamic (overall isotropic, $B = 2.0\,\text{Å}^2$)	0.614	0.828	0.818
(iv)	$(\sin\theta/\lambda)_{max} = 0.6\,\text{Å}^{-1}$	0.613	0.832	0.811
	Dynamic (individual anisotropic)	0.630	0.847	0.831
	Ab initio	0.972		

multipole models were pursued; a single data cutoff was employed $[(\sin\theta/\lambda)_{max} = 1.0\,\text{Å}^{-1}]$; the effects of random experimental errors and corresponding least-squares weights were not included; intermolecular interactions were ignored; all systematic experimental errors were ignored (*e.g.* extinction, thermal diffuse scattering). Neverthess, this work represents the first systematic study of the capabilities of the rigid pseudo-atom model to retrieve important molecular properties from X-ray diffraction data, its conclusions are directly relevant to experimental studies which share that objective and the results will serve as a benchmark for our future, more detailed, investigations. These will include the incorporation of intermolecular interactions and the exploration of improved models for the radial dependence of the multipole functions. We remain optimistic about the potential for accurate X-ray data to provide quantitative estimates of EFG's, molecular moments and even intermolecular interaction energies under ideal conditions, as there is little in the present general conclusions which suggest otherwise. Because many systems show widely differing behaviour (*e.g.* for acetylene the second moment tensor, and hence quadrupole moment, can be retrieved almost exactly, while for acetamide and HMT large variations in derived properties are observed), model studies of this kind would seem to be an extremely valuable adjunct to any experimental work aiming to extract these molecular properties.

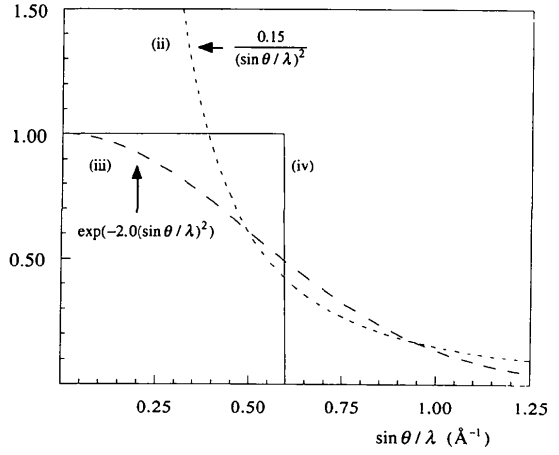

Fig. 9. Graphical representation of different weighting schemes (see text for details).

Professor R. Boese kindly supplied the positional and thermal parameters for borazine. We are grateful to the Australian Research Council for support of this work.

References

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

Boswell, M. T., Gore, S. D., Patil, G. P. & Taillie, C. (1993). In *Handbook of Statistics 9. Computational Statistics*, edited by C. R. Rao, pp. 661–721. Amsterdam: North Holland.

Brown, A. S. & Spackman, M. A. (1994). *Mol. Phys.* **83**, 551–566.

Bruning, H. & Feil, D. (1992). *Acta Cryst.* A**48**, 865–872.

Chandler, G. S. & Phillips, R. A. (1986). *J. Chem. Soc., Faraday Trans. 2*, **82**, 573–592.

Chandler, G. S. & Spackman, M. A. (1978). *Acta Cryst.* A**34**, 341–343.

Chandler, G. S., Figgis, B. N., Reynolds, P. A. & Wolff, S. K. (1994). *Chem. Phys. Lett.* **225**, 421–426.

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

Coppens, P. (1967). *Science*, **158**, 1577–1579.

Cromer, D. T. (1974). *International Tables for X-ray Crystallography*, Vol. IV, pp. 148–151. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

De Vries, R. Y., Briels, W. J. & Feil, D. (1994). *Acta Cryst.* A**50**, 383–391.

Dunning, T. H. Jr (1970). *J. Chem. Phys.* **53**, 2823–2833.

Feil, D. & Moss, G. (1983). *Acta Cryst.* A**39**, 14–21.

Gajdardziska-Josifovska, M., McCartney, M. R., De Ruijter, W. J., Smith, D. J., Weiss, J. K. & Zuo, J. M. (1993). *Ultramicroscopy*, **50**, 285–299.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

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.

Hinchliffe, A. (1987). *Ab Initio Determination of Molecular Properties*. Bristol: Adam Hilger.

Hirshfeld, F. L. (1977). *Theor. Chim. Acta*, **44**, 129–138.

Howard, S. T., Huke, J. P., Mallinson, P. R. & Frampton, C. S. (1994). *Phys Rev. B*, **49**, 7124–7136.

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

Howard, S. T., Hursthouse, M. B., Lehmann, C. W. & Poyner, E. A. (1995). *Acta Cryst.* B**51**, 328–337.

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.

Jeffrey, G. A., Ruble, J. R., McMullan, R. K. & Pople, J. A. (1987). *Proc. R. Soc. London Ser. A*, **414**, 47–57.

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

Klooster, W. T. (1992). PhD Thesis. University of Twente, Enschede, The Netherlands.

Krijn, M. P. C. M., Graafsma, H. & Feil, D. (1988). *Acta Cryst.* B**44**, 609–616.

Le Page, Y. & Gabe, E. J. (1979). *J. Appl. Cryst.* **12**, 464–466.

McMullan, R. K., Epstein, J., Ruble, J. R. & Craven, B. M. (1979). *Acta Cryst.* B**35**, 688–691.

McMullan, R. K., Kvick, Å. & Popelier, P. (1992). *Acta Cryst.* B**48**, 726–731.

Moss, G. & Coppens, P. (1980). *Chem Phys. Lett.* **75**, 298–302.

Moss, G. R., Souhassou, M., Blessing, R. H., Espinosa, E. & Lecomte, C. (1995). *Acta Cryst.* B**51**, 650–660.

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

O'Keeffe, M. & Spence, J. C. H. (1994). *Acta Cryst.* A**50**, 33–45.

Poorthuis, G. H. A. & Feil, D. (1994). *J. Mol. Struct.* **314**, 155–167.

Ritchie, J. P., Cromer, D. T., Stewart, R. F., Wasserman, H. J. & Ryan, R. R. (1985). Proceedings of 8th International Symposium on Detonation, S-046, Albuquerque, New Mexico.

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

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M. & Montgomery, J. A. (1993). *J. Comput. Chem.* **14**, 1347–1363.

Spackman, M. A. (1979). PhD Thesis. University of Western Australia.

Spackman, M. A. (1992). *Chem. Rev.* **92**, 1769–1797.

Spackman, M. A., Weber, H.-P. & Craven, B. M. (1988). *J. Am. Chem. Soc.* **110**, 775–782.

Stevens, E. D. (1978). *Acta Cryst.* B**34**, 544–551.

Stewart, R. F. (1969). *J. Chem. Phys.* **51**, 4569–4577.

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. & Jensen, L. H. (1967). *Acta Cryst.* **23**, 1102–1105.

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.

Swaminathan, S., Craven, B. M., Spackman, M. A. & Stewart, R. F. (1984). *Acta Cryst.* B**40**, 398–404.

Thakkar, A. J., Koga, T., Saito, M. & Hoffmeyer, R. E. (1993). *Int J. Quantum Chem. Quantum Chem. Symp.* **27**, 343–354.

Velders, G. J. M. & Feil, D. (1989). *Acta Cryst.* B**45**, 359–364.

Wolff, S. K. (1995). PhD Thesis. University of Western Australia.