

# Fast estimation of crystal densities

**Detlef Walter Maria Hofmann**

Department of Computer Science, University of  
Bonn, Römerstrasse 164, 53117 Bonn, Germany

Correspondence e-mail:  
detlef.hofmann@gmd.de

A simple, efficient and accurate method for the estimation of crystal densities is of interest for different applications. By analysis of the Cambridge Structural Database (CSD) we derived an average atom volume method to estimate the cell volume for a given formula and  $Z$  value. This method extends the work of Mighell *et al.* (1987) by inclusion of the thermal expansion and error estimation.

Received 23 July 2001

Accepted 20 December 2001

## 1. Introduction

It is often useful to know what the density of an organic or metal-organic crystal should be, particularly in evaluating the correctness of a chemical composition and/or experimental unit-cell parameters. The former 18 Å<sup>3</sup> rule (Kempster & Lipson, 1972) has been improved by replacing the average volume of the atoms by individual volumes for each element (Mighell *et al.*, 1987). The presented method extends this approach to the effect of temperature and derives an error estimation for the average volumes.

In some industrial applications a certain density of the crystal may be required. In the process of crystal engineering a simple method of density prediction can exclude unprofitable candidates at a very early stage. An accurate density estimation has been developed for the purpose of propellant optimization (Ammon & Mitchell, 1998; Stine, 1981). However, these methods have been parameterized for a few elements and the density is not as simple to calculate as it is by the presented method.

## 2. Method

To derive actual parameters for density prediction, the CSD (Release April 2001; Allen & Kennard, 1993) was analyzed. Structures with disorder were not taken into account. In a first step, 4815 unusable structures, for our purpose, were screened off. This concerns structures with a given  $Z$  value of zero, structures with a cell volume of zero, structures without a given cell and structures with questionable cell content. Sometimes the cell content contains zero as the frequency for an element, *e.g.* C1 H1 Cl1 Fe1 N0 O1 for the structure

**Table 1**  
Characteristic values during the iterations of sample 1.

Iteration	$\alpha_{jl} * 10^5$	Retained structures	Intercept $a$	Slope $b * 10^5$
0	0.000	9112	1.000	0.000
1	0.884	4474	0.970	0.857
2	0.964	5406	1.003	0.081
3	1.026	5689	0.997	0.062
4	1.030	5889	0.999	0.004
5	1.006	6013	1.001	-0.025
6	0.978	6039	1.001	-0.028
7	0.948	6071	1.001	-0.029
8	0.924	6043	1.000	-0.025
9	0.873	6040	1.001	-0.051
10	0.876	6036	1.000	0.003
11	0.866	6031	1.000	-0.010
12	0.867	6041	1.000	0.000
13	0.876	6052	1.000	0.010
14	0.885	6050	1.000	0.008
15	0.892	6044	1.000	0.007
16	0.899	6044	1.000	0.007
17	0.906	6045	1.000	0.007
18	0.909	6050	1.000	0.004
19	0.916	6050	1.000	0.006
20	0.918	6049	1.000	0.002

ABYXCZ. Application of the screens left 182 239 structures, which we used to derive average volumes for the elements.<sup>1</sup>

The volume of a crystal was presumed to be the sum of the average volumes  $\bar{v}_i$  of the elements in the unit cell, where  $n_i$  is the number of atoms of element  $i$  in the unit cell. Assuming that the volume varies linearly with the temperature, we have

$$V_{\text{est}} = \sum_{i=1}^{100} n_i \bar{v}_i (1 + \bar{\alpha} T) = \mathbf{n} \bar{\mathbf{v}} (1 + \bar{\alpha} T), \quad (1)$$

where  $\bar{\alpha}$  is the average thermal expansion coefficient. The sum runs over all elements from hydrogen (atom number 1) to fermium (atom number 100). The database contains structures up to americium and for the higher atom numbers no average volumes can be derived with this method.

In order to estimate the parameters  $\bar{v}_i$  and  $\bar{\alpha}$  and their accuracies, the total number of structures was divided into 20 samples  $j$  of 9112 structures. For each of these samples, individual average volumes  $v_{ij}$  and thermal expansion coefficients  $\alpha_j$  were determined.

$$\begin{aligned} \mathbf{n}_{j1} \mathbf{v}_{j1} (1 + \alpha_{j1} T) &= V_{\text{obs},j1} \\ &\dots = \dots \\ \mathbf{n}_{jk} \mathbf{v}_{j1} (1 + \alpha_{j1} T) &= V_{\text{obs},jk} \\ &\dots = \dots \\ \mathbf{n}_{j9112} \mathbf{v}_{j1} (1 + \alpha_{j1} T) &= V_{\text{obs},j9112}. \end{aligned} \quad (2)$$

To obtain the values  $v_{ij}$  and  $\alpha_j$  the overdetermined non-linear equation system (2) was solved iteratively. The parameters  $v_{ij}$  or  $\alpha_{jl}$  were kept constant alternately, whereby the iteration is indicated by the index  $l$ . In a first approximation the above equation was solved for each sample, while

<sup>1</sup>Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK0101). Services for accessing these data are described at the back of the journal.

neglecting temperature dependence in order to obtain a first approximation  $\mathbf{v}_{j0}$ .

$$\mathbf{n}_j \mathbf{v}_{j0} (1 + \alpha_{jl} T) = \mathbf{V}_{\text{obs},j} \text{ with } \alpha_{j0} = 0. \quad (3)$$

This is a multilinear regression problem that can be solved by a least-squares procedure. For this purpose we use the singular value decomposition for overdetermined systems (Press *et al.*, 1992; Stoer & Bulirsch, 1980; Golub & Van Loan, 1989), which is a general method for solving an overdetermined linear equation system. The algorithm is based on the theorem that the matrix  $\mathbf{n}$  can be expressed by the two orthonormal column matrices  $\mathbf{A}$  and  $\mathbf{B}$  and the positive definite diagonal matrix  $\mathbf{w}$  in the following way

$$\mathbf{n}_j = \mathbf{A} \mathbf{w} \mathbf{B}^T. \quad (4)$$

The solution of the unknown parameters  $\mathbf{v}_{j0}$  is then given by

$$\mathbf{v}_{j0} = \mathbf{B} \mathbf{w}^{-1} \mathbf{A}^T \mathbf{V}_{\text{obs},j}. \quad (5)$$

If  $v_{ijl}$  are taken as fixed, the equation system becomes a linear regression problem. For each structure  $k$  out of the sample, the ratio  $V_{\text{obs},jk}/V_{\text{est},jkl}$  versus the temperature can be plotted as

$$V_{\text{obs},jk}/V_{\text{est},jkl} = a + bT. \quad (6)$$

From the slope  $b$  and the intercept  $a$  of the regression, improved values for the thermal expansion coefficients and the average volumes can be calculated as

$$\begin{aligned} V_{\text{obs},jk} &= V_{\text{est},jkl} (a + bT) \\ &= \mathbf{n} \mathbf{v}_{j0} c_l (1 + \alpha_{jl} T) (a + bT) \\ &= \mathbf{n} \mathbf{v}_{j0} c_l a (1 + \alpha_{jl} T + b/aT + \alpha_{jl} b/aT^2) \\ &\simeq \mathbf{n} \mathbf{v}_{j0} c_l a (1 + \alpha_{jl} T + b/aT) \\ &= \mathbf{n} \mathbf{v}_{j0} c_{l+1} (1 + \alpha_{j,l+1} T) \\ &= \mathbf{n} \mathbf{v}_{j,l+1} (1 + \alpha_{j,l+1} T) \text{ with } c_0 = 1. \end{aligned} \quad (7)$$

In the first round all 9112 structures of a particular set were used and the starting value for  $c_0$  was set to one. Subsequently, only structures were retained whose experimental volume  $V_{\text{obs}}$  deviated by less than 5% from the estimated volume  $V_{\text{est}}$ . Organic and metal-organic structures with values below this threshold are supposed to include undiscovered solvent (Stalick, 2001).

$$1/1.05 < V_{\text{est}}/V_{\text{obs}} < 1.05. \quad (8)$$

After the screening we returned back to (3). The procedure was repeated until self-consistency of the expansion coefficient and the atomic volumes were achieved. In Table 1 we give the number of retained structures, the intercept  $a$ , the slope  $b$  and the thermal expansion coefficients  $\alpha_{l+1}$  as a function of the iteration for sample 1. The average volume  $\bar{v}_i$  of an element  $i$  is the arithmetic mean of the ten volumes determined for the different samples.

$$\bar{v}_i = (1/20) \sum_{j=1}^{20} v_{ij}. \quad (9)$$

The standard deviation  $\sigma$  and the mean error  $\Delta v_i$  for an average volume is given by

**Table 2**

The average volume  $\bar{v}_i$  of elements at 298 K.

Atom number	Element	$\bar{v}$ (Å <sup>3</sup> )	$\Delta v$ (Å <sup>3</sup> )	Mighell <i>et al.</i> (1987)	Difference (%)
1	H	5.08	0.04	5.40	-5
2	He	-	-	10.00	-
3	Li	22.6	0.9	18.00	25
4	Be	36	4	24.50	45
5	B	13.24	0.17	13.00	1
6	C	13.87	0.05	13.94	0
7	N	11.8	0.3	11.00	7
8	O	11.39	0.17	10.00	13
9	F	11.17	0.15	11.50	-2
10	Ne	-	-	20.00	-
11	Na	26	3	18.00	44
12	Mg	36	4	38.00	-4
13	Al	39.6	1.3	34.50	14
14	Si	37.3	0.3	39.00	-4
15	P	29.5	0.2	30.00	-1
16	S	25.2	0.3	26.00	-2
17	Cl	25.8	0.3	25.00	3
18	Ar	-	-	30.00	-
19	K	36	3	31.00	15
20	Ca	45	6	41.00	9
21	Sc	42	6	21.00	99
22	Ti	27.3	1.8	25.50	7
23	V	24.0	1.8	39.00	-38
24	Cr	28.1	1.5	34.00	-17
25	Mn	31.9	0.9	31.50	1
26	Fe	30.4	0.7	29.00	4
27	Co	29.4	1.3	27.00	8
28	Ni	26	2	26.50	-1
29	Cu	26.9	1.0	24.00	12
30	Zn	39	3	35.00	12
31	Ga	37.8	1.5	39.00	-3
32	Ge	41.6	1.3	39.00	6
33	As	36.4	1.3	36.50	0
34	Se	30.3	1.1	28.50	6
35	Br	32.7	0.6	30.50	7
36	Kr	-	-	40.00	-
37	Rb	42	5	39.00	8
38	Sr	47	4	39.00	21
39	Y	44	3	37.00	17
40	Zr	27	2	40.00	-31
41	Nb	37	2	35.50	3
42	Mo	38	2	37.00	2
43	Tc	38	5	37.00	2
44	Ru	37.3	0.9	43.00	-13
45	Rh	31.2	1.0	39.00	-20
46	Pd	35	3	34.50	2
47	Ag	35	2	29.50	18
48	Cd	51	3	35.00	45
49	In	55	3	64.00	-14
50	Sn	52.8	1.3	49.00	7
51	Sb	48.0	1.6	46.00	4
52	Te	46.7	1.9	48.00	-2
53	I	46.2	0.4	44.00	5
54	Xe	45	8	50.00	-9
55	Cs	46	2	53.00	-14
56	Ba	66	4	41.00	61
57	La	58	5	65.00	-10
58	Ce	54	5	60.00	-9
59	Pr	57	7	60.00	-4
60	Nd	50	4	60.00	-16
61	Pm	-	-	55.00	-
62	Sm	50	4	55.00	-9
63	Eu	53	4	55.00	-4
64	Gd	56	5	50.00	12
65	Tb	45	7	50.00	-10
66	Dy	50	7	50.00	0
67	Ho	42	5	50.00	-15
68	Er	54	5	50.00	8
69	Tm	49	6	45.00	8
70	Yb	59	4	45.00	30

**Table 2 (continued)**

Atom number	Element	$\bar{v}$ (Å <sup>3</sup> )	$\Delta v$ (Å <sup>3</sup> )	Mighell <i>et al.</i> (1987)	Difference (%)
71	Lu	35	4	45.00	-21
72	Hf	40	4	45.00	-10
73	Ta	43	2	43.00	0
74	W	38.8	1.6	46.00	-15
75	Re	42.7	1.8	51.00	-16
76	Os	41.9	0.6	43.00	-2
77	Ir	34.3	0.9	29.50	16
78	Pt	38	2	42.00	-9
79	Au	43	2	31.00	39
80	Hg	38.0	1.4	38.00	0
81	Tl	54	4	58.00	-6
82	Pb	52	4	43.00	21
83	Bi	60	4	49.00	22
84	Po	-	-	50.00	-
85	At	-	-	55.00	-
86	Rn	-	-	60.00	-
87	Fr	-	-	70.00	-
88	Ra	-	-	60.00	-
89	Ac	74	-	75.00	-
90	Th	56	5	85.00	-33
91	Pa	60	30	80.00	-25
92	U	58	4	61.00	-5
93	Np	45	6	70.00	-35
94	Pu	-	-	70.00	-
95	Am	17	-	70.00	-
96	Cm	-	-	70.00	-
97	Bk	-	-	70.00	-
98	Cf	-	-	70.00	-
99	Es	-	-	70.00	-
100	Fm	-	-	70.00	-

$$\sigma(\bar{v}_i) = \left[ (1/19) \sum_{j=1}^{20} (v_{ij} - \bar{v}_i)^2 \right]^{1/2} \quad (10)$$

$$\Delta v_i = 1.03/(20)^{1/2} \sigma(\bar{v}_i). \quad (11)$$

The average thermal expansion coefficient  $\bar{\alpha}$  and its mean error  $\Delta\alpha$  was calculated to be analogous.

The density  $\rho$  of a crystal can be estimated from the average volumes and the atom mass  $m_i$  obtained by

$$\rho = \left( \sum_{i=1}^{100} m_i n_i \right) / \left[ \sum_{i=1}^{100} n_i \bar{v}_i (1 + \bar{\alpha} T) \right]. \quad (12)$$

The mean error for the density  $\Delta\rho$  can be calculated from the determined mean errors of the average volumes  $\Delta v_i$  and the thermal expansion coefficient  $\Delta\alpha$  by

$$\Delta\rho = \pm\rho \left[ \left( \frac{\sum_{i=1}^{100} n_i \Delta v_i}{\sum_{i=1}^{100} n_i \bar{v}_i} \right)^2 + \left( \frac{T \Delta\alpha}{1 + \bar{\alpha} T} \right)^2 \right]^{1/2}. \quad (13)$$

### 3. Results

The coefficient of the thermal expansion was determined to be  $(0.95 \pm 0.03) \times 10^{-4} \text{ K}^{-1}$ . The values of the average volumes

**Table 3**  
Structures for which the volume is more than 25% overestimated ( $V_{\text{obs}}/V_{\text{est}} < 0.8$ ).

Reference code	Compound	$V_{\text{obs}}/V_{\text{est}}$
BAFORM	Barium formate	0.775831
BIFVAN	Dilithium malonate	0.715611
GAFORM01	Gadolinium formate	0.747237
TERMOS	Lithium phosphonoacetate	0.796253
ZZZVZO01	Dirubidium tartrate	0.798727

$\bar{v}_i$  of the elements at room temperature (298 K) with their mean errors are contained in Table 2.

For the elements actinium and americum no standard derivation can be calculated. For both cases only a single compound containing these elements occurs in the database. As just two compounds with protactinium are available, the largest mean error is calculated for this element. The largest difference (99%) between the average volumes in this work and the former work (Mighell *et al.*, 1987) is observed for scandium. This element occurs in 95 structures and the difference can be explained by different oxidation states  $\text{Sc}^0$  and  $\text{Sc}^{3+}$ . Recent scandium complexes contain more  $\text{Sc}^0$  than previous complexes. The different oxidations states for this element and other metals cause large mean errors  $\Delta v_i$ , even if they occur frequently in the database. Similar to the metals, the mean error of the halides depends on the different chemical environment, bonded or ionic, and not on the frequency of the element in the database, *e.g.* the volume of iodine can be calculated from the covalent radius 2.20 Å to 44.6 Å<sup>3</sup> and from the ionic radius 2.15 Å to 41.6 Å<sup>3</sup> (Weast, 1987), which is reflected by the large variation  $\Delta v_I$ .

We observe 1746 structures with a per cent difference in the volume above 25%. Three different reasons for these unusual deviations were found:

(i) The volume is over- or underestimated by more than 75% in 737 cases, often by an integer multiple ( $V_{\text{obs}} \approx nV_{\text{est}}$  resp.  $V_{\text{obs}} \approx V_{\text{est}}/n$ ). The reason is mainly that the  $Z$  value has been entered incorrectly into the database. Earlier versions of

the CSD checking software did not include a search for void spaces when  $Z$  is too low. This has now been implemented.

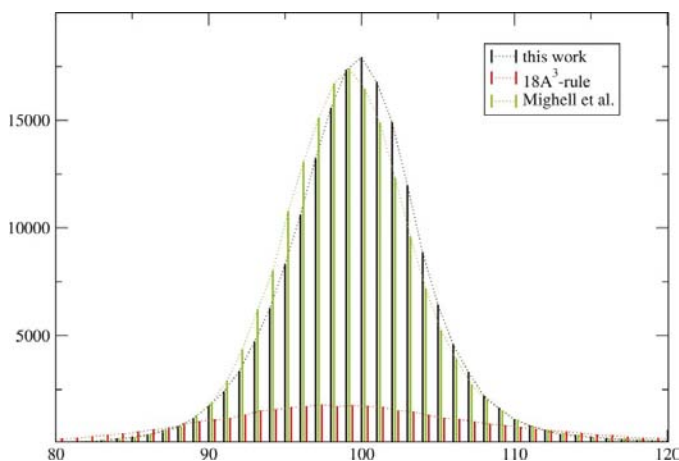
(ii) For 513 compounds the cell volume is underestimated by 25–75%. Visual inspection did not indicate any obvious common feature. A possible reason may be the unrecognized or disordered inclusion of several solvent molecules during crystallization.

(iii) The volume is overestimated for 25–75% in 496 cases. These compounds are often highly ionic and furthermore the cells are very small. A few examples are given in Table 3. This indicates that the average volumes determined may not be applicable to inorganic or mainly inorganic substances. To handle this problem additional sets of average volumes for inorganics and for metals and intermetallics were introduced by Stalick (1984).

Fig. 1 shows a histogram which illustrates the ratio between the observed and estimated volumes  $V_{\text{obs}}/V_{\text{est}}$  calculated with the derived values, according to previous studies (Mighell *et al.*, 1987), and using the 18 Å<sup>3</sup> rule. Comparison of the three approximations shows that the introduction of individual average volumes (Mighell *et al.*, 1987) significantly increased the prediction of densities. The standard deviation of the 18 Å<sup>3</sup> rule (Kempster & Lipson, 1972) in Fig. 1 improved from 9.04 to 4.13. Although the standard deviation of 4.00 in this work is a slight improvement, the inclusion of the thermal expansion mainly corrects the average value for prediction. The former method (Mighell *et al.*, 1987) slightly underestimates the average volume by 1.2% (the 18 Å<sup>3</sup> rule by 1.7%). This error is reduced to 0.3% in the present work. The new method in addition allows the calculation of the mean error for the predicted density.

An improvement of the presented estimation can be achieved by additional terms to (1). To take into account the local environment of a certain atom one can introduce terms (Ammon & Mitchell, 1998) that depend on functional groups. However, the functional-group-dependent method is restricted to C, H, N, O and F, and the density estimation is more complicated and is based on significantly more parameters than the simple method presented here.

We would like to thank W. D. S. Motherwell for the support in all questions concerning the CSD and the National Institute of Standards and Technology (USA) for providing the program NBS\*AIDS83. For the financial support we thank the 'Deutsche Forschungsgemeinschaft' (Proj. LE 491/12–1).



**Figure 1**  
Ratio of estimated and experimental volume  $V_{\text{obs}}/V_{\text{est}}$  for all 182 239 structures within  $1 \pm 0.20$ .

## References

- Allen, F. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.  
 Ammon, H. L. & Mitchell, S. (1998). *Propellants Explos. Pyrotech.* **23**, 260–265.  
 Golub, G. H. & Van Loan, C. (1989). *Matrix Computations*, ch. 5.2.6, 2nd ed. Baltimore: John Hopkins University Press.  
 Kempster, C. J. E. & Lipson, H. (1972). *Acta Cryst.* **B28**, 3674.  
 Mighell, A. D., Hubbard, C. R., Stalick, J. K., Santoro, A., Snyder, R. L., Holomany, M., Seidel, J. & Lederman, S. (1987). National Bureau of Standards, Gaithersburg, MD 20899, USA.

- Press, H. W., Teukolsky, S., Vetterling, W. T. & Flannery, B. P. (1992). *Singular Value Decomposition*, pp. 51–63, 2nd ed. Cambridge University Press.
- Stalick, J. K. (1984). Technical Note 1190, pp. 52–55. National Bureau of Standards, USA.
- Stalick, J. K. (2001). Personal Communication.
- Stine, J. R. (1981). Report LA-8920. Los Alamos National Laboratory, New Mexico, USA.
- Stoer, J. & Bulirsch, R. (1980). *Introduction to Numerical Analysis*, ch. 6.7. New York: Springer Verlag.
- Weast, R. (1987). *CRC Handbook of Chemistry and Physics*, Student ed., 1st ed. Boca Raton, Florida: CRC Press, Inc.