

- BOYSEN, H. (1990). In *Phase Transitions in Ferroelastic and Co-elastic Crystals*, edited by E. K. H. SALJE, pp. 320-335. Cambridge Univ. Press.
- BRUCE, A. D. & COWLEY, R. A. (1981). *Structural Phase Transitions*. London: Taylor & Francis.
- CHAUDHARI, P. (1987). *Jpn. J. Appl. Phys.* **26**, 2023-2034.
- CHENG, C., HEINE, V. & JONES, I. L. (1990). *J. Phys. Condens. Matter*, **2**, 5097-5114.
- CROSS, L. E. (1987). *Ferroelectrics*, **76**, 241-267.
- DARLINGTON, C. N. W. (1990). In *Phase Transitions in Ferroelastic and Co-elastic Crystals*, edited by E. K. H. SALJE, pp. 270-282. Cambridge Univ. Press.
- DATTAGUPTA, S., HEINE, V., MARAIS, S. & SALJE, E. (1991). *J. Phys. Condens. Matter*. In the press.
- DEUTSCHER, G. & MÜLLER, K. A. (1987). *Phys. Rev. Lett.* **59**, 1745-1747.
- DEVARAJAN, V. & SALJE, E. (1984). *J. Phys. C*, **17**, 5525-5537.
- DOVE, M. T. (1990). In *Phase Transitions in Ferroelastic and Co-elastic Crystals*, edited by E. K. H. SALJE, pp. 297-311. Cambridge Univ. Press.
- FLOCKEN, J. W., GUENTHER, R. A., HARDY, J. R. & BOYER, L. L. (1985). *Phys. Rev. B*, **31**, 7252-7260.
- GUFAN, YU. M. & LARIN, E. S. (1980). *Sov. Phys. Solid State*, **22**, 270-273.
- HARRIS, M. J., SALJE, E. K. H. & GÜTTLER, B. K. (1990). *J. Phys. Condens. Matter*, **2**, 5517-5527.
- HATCH, D. M., ARTMAN, J. I. & BOERIO-GOATES, J. (1990). *Phys. Chem. Miner.* **17**, 334-343.
- IGUCHI, E. (1991). *Solid State Chem.* In the press.
- IMRY, Y. (1975). *J. Phys. C*, **8**, 567-577.
- International Tables for X-ray Crystallography* (1969). Vol. I. Birmingham: Kynoch Press.
- ISHIBASHI, Y., HARA, K. & SAWADA, A. (1988). *Physica (Utrecht)*, **B150**, 258-264.
- JORGENSEN, J. D., VEAL, B. W., KWOK, W. K., CRABTREE, G. W., UMEZAWA, A., NOWICKI, L. J. & PAULIKAS, A. P. (1987). *Phys. Rev. B*, **36**, 5731-5734.
- LUTI, B. & REHWALD, W. (1981). *Curr. Phys.* **23**, 131-184.
- LYNDEN-BELL, R. M., FERRARIO, M., McDONALD, I. R. & SALJE, E. (1989). *J. Phys. Condens. Matter*, **1**, 6523-6542.
- MARAIS, S., HEINE, V. & SALJE, E. (1991). In preparation.
- MOTT, N. F. (1990). *Adv. Phys.* **39**, 55-81.
- OLESKY, C. & PRYSZTAWA, J. (1983). *Physica (Utrecht)*, **121A**, 145-149.
- PALMER, D. C., SALJE, E. K. H. & SCHMAHL, W. W. (1989). *Phys. Chem. Miner.* **16**, 714-719.
- PARLINSKI, K. (1988). *Comp. Phys. Rep.* **8**, 157.
- REDFERN, S. A. T., GRAEME-BARBER, A. & SALJE, E. (1988). *Phys. Chem. Miner.* **16**, 157-163.
- REDFERN, S. A. T. & SALJE, E. (1987). *Phys. Chem. Miner.* **14**, 189-195.
- REDFERN, S. A. T. & SALJE, E. (1988). *J. Phys. C*, **21**, 277-285.
- ROCQUET, P. & COUZI, M. (1985). *J. Phys. C*, **18**, 6571-6579.
- SALJE, E. (1985). *Phys. Chem. Miner.* **12**, 93-98.
- SALJE, E. (1988). *Phys. Chem. Miner.* **15**, 336-348.
- SALJE, E. (1989). *Philos. Mag. Lett.* **59**, 219-221.
- SALJE, E. (1989). *Philos. Trans. R. Soc. London Ser. A*, **328**, 409-416.
- SALJE, E. K. H. (1990a). Editor. *Phase Transitions in Ferroelastic and Co-elastic Crystals*. Cambridge Univ. Press.
- SALJE, E. (1990b). *Philos. Mag. Lett.* **62**, 277-282.
- SALJE, E. & BISMAYER, U. (1989). *J. Phys. Condens. Matter*, **1**, 6967-6976.
- SALJE, E., BISMAYER, U. & JANSEN, M. (1987). *J. Phys. C*, **20**, 3613-3620.
- SALJE, E. & DEVARAJAN, V. (1986). *Phase Transitions*, **6**, 235-248.
- SALJE, E., KUSCHOLKE, B. & WRUCK, B. (1985). *Phys. Chem. Miner.* **12**, 132-140.
- SALJE, E., KUSCHOLKE, B., WRUCK, B. & KROLL, H. (1985). *Phys. Chem. Miner.* **12**, 99-107.
- SALJE, E. & PARLINSKI, K. (1991). *Superconductor Science and Technology*. In the press.
- SALJE, E. & WRUCK, B. (1988). *Phys. Chem. Miner.* **16**, 140-147.
- SALJE, E. K. H., WRUCK, B. & THOMAS, H. (1991). *Z. Phys.* **B82**, 399-404.
- SAPRIEL, J. (1975). *Phys. Rev. B*, **12**, 5128-5140.
- SCHMAHL, W. W. & SALJE, E. (1989). *Phys. Chem. Miner.* **16**, 790-796.
- SCHMAHL, W. W., SALJE, E. & LIANG, W. Y. (1988). *Philos. Mag. Lett.* **58**, 173-181.
- STOKES, H. T. & HATCH, D. M. (1988). *Isotropy Subgroups of the 230 Crystallographic Space Groups*. Singapore: World Scientific.
- SUZUKI, T. & ISHIBASHI, Y. (1987). *J. Phys. Soc. Jpn.* **56**, 596-602.
- TOLEDANO, J. C. (1979). *Phys. Rev. B*, **20**, 1147-1156.
- TOLEDANO, J. C. & TOLEDANO, P. (1980). *Phys. Rev. B*, **21**, 1139-1172.
- TOLEDANO, J. C. & TOLEDANO, P. (1982). *Phys. Rev. B*, **25**, 1946-1964.
- WADHAWAN, V. K. (1982). *Phase Transitions*, **3**, 3-103.
- WINKLER, B., DOVE, M. T., SALJE, E. K. H., LESLIE, M. & PALOSZ, B. (1991). *J. Phys. Condens. Matter*. Submitted.
- WRUCK, B., SALJE, E. & GRAEME-BARBER, A. (1991). *Phys. Chem. Miner.* In the press.
- YOSHIHARA, A., YOSHIZAWA, M., YASUDA, H. & FUJIMURA, T. (1985). *Jpn. J. Appl. Phys. Suppl.* **24**, 367-369.

Acta Cryst. (1991). **A47**, 469-480

Derivation of Exponential Joint Probability Distributions of Structure Factors in *P1* with Maximum Entropy and Irreducible Cluster Integrals

BY M. J. KRONENBURG, R. PESCHAR AND H. SCHENK

Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

(Received 10 October 1990; accepted 7 May 1991)

Abstract

Recently the application of the maximum-entropy method to direct methods has been initiated for

a priori uniformly and independently distributed atoms, introducing non-uniformity in direct space by putting constraints on the expected values of the distribution [Bricogne (1984). *Acta Cryst.* **A40**, 410-

445; (1988). *Acta Cryst. A* **44**, 517-545]. In this paper a start is made in using the maximum-entropy principle for deriving exponential joint probability distributions of structure factors for a chemically more realistic model of *a priori* non-uniformly and non-independently distributed atoms. The maximum-entropy equations are obtained by treating the atomic positions as well as the reciprocal vectors as random variables and applying constraints on the maximum of the distribution. The interdependence of the Lagrange multipliers leads to inequalities which may be compared with the Karle-Hauptman inequalities. The radial interatomic correlations such as minimal interatomic distances lead to integrals whose evaluation *via* the cluster integral mechanism is shown to be equivalent to those of the classical hard-sphere gas in an external field [Van Kampen (1961). *Physica (Utrecht)*, **27**, 783-792]. The Debye scattering equation results from these calculations. The exponential multiplet terms are expressed as cluster integrals. From the distribution of the single structure factor the influence of the interatomic correlations on the normalization procedure is assessed. The exponential triplet distribution up to order $N^{-3/2}$ is derived and is shown to be in agreement with the exponential Edgeworth result [Karle & Gilardi (1973). *Acta Cryst. A* **29**, 401-407]. The effect of the radial interatomic correlations on the triplet distribution is discussed. The exponential quartet distribution up to order N^{-1} is derived, and found to be equal to the well known result [Hauptman (1975). *Acta Cryst. A* **31**, 617-679, 680-687] except for some normalization terms resulting from the interatomic correlations.

1. Introduction

The derivation of joint probability distributions (j.p.d.'s) using the characteristic function (Karle & Hauptman, 1953; Hauptman & Karle, 1958; Bertaut, 1955, 1956; Klug, 1958; Naya, Nitta & Oda, 1965; Karle, 1972; Hauptman, 1975*a, b*, 1976; Peschar & Schenk, 1987; and many other papers) may be understood using the Fourier integral representation of the Dirac delta function in the complex plane ($z \in \mathbb{C}$):

$$\delta(z) = (2\pi)^{-2} \int_x \exp(ix \cdot z) dx \quad (1.1)$$

where the integral is over the complex plane and the dot product in the complex plane is defined as

$$a \cdot b \equiv \text{Re}(a) \text{Re}(b) + \text{Im}(a) \text{Im}(b). \quad (1.2)$$

Having defined the complex structure factors in a certain space group $F(\mathbf{k}_\mu, \mathbf{r}_\nu)$ as functions of the atomic positions in the asymmetric unit of the cell \mathbf{r}_ν ($\nu = 1, \dots, N$) and reciprocal vectors \mathbf{k}_μ ($\mu = 1, \dots, M$), (1.1) can be used to evaluate the joint probability distribution of the set of M corresponding

complex structure-factor variables F_μ :

$$P(F_\mu) = \int_{\mathbf{r}_\nu} \cdots \int_{\mu} [\delta[F_\mu - F(\mathbf{k}_\mu, \mathbf{r}_\nu)] P(\mathbf{r}_\nu) d\mathbf{r}_\nu. \quad (1.3)$$

$P(\mathbf{r}_\nu)$ is the *a priori* probability distribution of the atomic positions in the unit cell. With (1.1) this becomes the expression with the characteristic function:

$$P(F_\mu) = (2\pi)^{-2M} \int_{x_\mu} \cdots \int \exp\left(-i \sum_{\mu} x_\mu \cdot F_\mu\right) \\ \times \int_{\mathbf{r}_\nu} \cdots \int \exp\left[i \sum_{\mu} x_\mu \cdot F(\mathbf{k}_\mu, \mathbf{r}_\nu)\right] \\ \times P(\mathbf{r}_\nu) d\mathbf{r}_\nu dx_\mu. \quad (1.4)$$

After Taylor expansion of the first integrand and moment-cumulant transformation and a change of variables where phase-restricted structure factors are concerned, this yields the Edgeworth series expansion in Hermite and/or Laguerre polynomials with the $|E|$ values and the phase invariants as arguments (Klug, 1958; Naya, Nitta & Oda, 1965; Peschar & Schenk, 1987). A final transformation may yield an exponential distribution (Karle, 1972; Karle & Gilardi, 1973).

The Karle-Hauptman inequalities bound the values of the set of M structure factors F_μ in (1.4) (Karle & Hauptman, 1950). As $P(F_\mu)$ must be zero outside these bounds, the polynomial expansion should yield an infinite number of roots there, which cannot be achieved with a limited number of expansion terms taken into account. In agreement with this, computations show that the series expansion does not converge uniformly on its domain, which justifies an attempt to find an alternative for (1.4).

Because of the delta function in (1.3), only contributions for which $F(\mathbf{k}_\mu, \mathbf{r}_\nu) = F_\mu$ are taken into account in (1.4), which may be compared with the micro-canonical integral of a classical gas with constant volume. The role of the Hamiltonian is played by the structure factor, being a function of the atomic coordinates which vary randomly through the unit cell (Jaynes, 1957). In analogy with the classical gas theory, application of maximum entropy to the hypersurface $F(\mathbf{k}_\mu, \mathbf{r}_\nu) = F_\mu$ yields for each structure factor F_μ a Lagrange multiplier β_μ which is a function of the observed intensity and possibly other invariants. It has already been demonstrated (Bricogne, 1984, 1988) that maximum entropy yields results that are useful for direct methods and that are consistent with evaluation of (1.4) with the saddle-point approximation. Here β_μ instead of λ_μ are used, since the maximum-entropy constraint is different from that used by Bricogne; the constraint is chosen to be the maximum instead of the expected value of the distribution.

In a previous paper (Kronenburg, Peschar & Schenk, 1990), the former concept was introduced to derive the j.p.d. for a single structure factor in P_1 employing a uniform *a priori* distribution of the atoms. In the present paper, however, the concept of random-variable reciprocal vectors is combined with the chemically more realistic model of non-overlapping hard spheres. This gives rise to irreducible cluster integrals (Mayer & Mayer, 1940) which are shown to be similar to the integrals arising from the partition function of a hard-sphere gas in an external field (Van Kampen, 1961). Application of this formalism to the j.p.d. for the triplet and the quartet phase sums leads to results which are largely consistent with the Edgeworth-series result. In particular, the radial interatomic correlations, responsible for the fluctuation in the Wilson curve, lead to conclusions about the normalization of $|E|^{\text{obs}}$ values for phasing procedures.

2. Maximum entropy in reciprocal space in P_1

The structure factors $F(\mathbf{k}_\mu, \mathbf{r}_\nu)$ are functions of both \mathbf{k}_μ and \mathbf{r}_ν , the former being a set of M vectors in real three-dimensional (reciprocal) space, and the latter being a set of N atomic positions in the real three-dimensional (direct) space of the unit cell:

$$F(\mathbf{k}_\mu, \mathbf{r}_\nu): \mathbb{R}^{3M} \otimes \mathbb{R}^{3N} \rightarrow \mathbb{C}^M. \quad (2.1)$$

Both \mathbf{k}_μ and \mathbf{r}_ν are random variables here and the \mathbf{k}_μ will be fixed to the measured reflections later. Applying the central limit theorem for both and taking the *a priori* probability distribution of the \mathbf{r}_ν to be $P(\mathbf{r}_\nu)$, one can safely assume that the probability distribution $P(\mathbf{k}_\mu, \mathbf{r}_\nu | F_\mu)$ is a Gaussian in the complex plane:

$$P(\mathbf{k}_\mu, \mathbf{r}_\nu | F_\mu) \propto \exp \left[-\sum_{\mu} \beta_{\mu} |F_{\mu} - F(\mathbf{k}_{\mu}, \mathbf{r}_{\nu})|^2 \right] P(\mathbf{r}_{\nu}). \quad (2.2)$$

With the assumption that there is a set of Lagrange multipliers that maximizes probability on the hypersurface $F(\mathbf{k}_\mu, \mathbf{r}_\nu) = F_\mu$, these multipliers are equivalent with the β_μ , which follows from results below. This justifies the use of (2.2) and the term maximum entropy attached to it; no effort is made here to prove the general mathematical equivalence of the central limit theorem with the maximum-entropy formalism. By integration over the atomic positions in order to arrive at a distribution in reciprocal space, $P(\mathbf{k}_\mu | F_\mu)$ becomes

$$P(\mathbf{k}_\mu | F_\mu) \propto \int \cdots \int_{\mathbf{r}_\nu} \exp \left[-\sum_{\mu} \beta_{\mu} |F_{\mu} - F(\mathbf{k}_{\mu}, \mathbf{r}_{\nu})|^2 \right] P(\mathbf{r}_{\nu}) d\mathbf{r}_{\nu}. \quad (2.3)$$

The β_μ are given by fixing the maximum in the complex plane at $|F_\mu| = |F_\mu|^{\text{obs}}$, which yields the set

of maximum-entropy (ME) equations:

$$d \log P(\mathbf{k}_\mu | F_\mu) / d|F_\mu| = 0 \quad \text{at } |F_\mu| = |F_\mu|^{\text{obs}}. \quad (2.4)$$

If the distribution is assumed to be symmetric in the $|F_\mu|$ around some point, these ME equations result in Lagrange multipliers β_μ equivalent to the λ_μ derived by using expected values (Bricogne, 1984, ME3), because in that case the maximum coincides with the expected values. When the resulting β_μ values are functions of both the constraint values and the random variables in \mathbb{C}^M , the random variables are averaged out. Substitution of these β_μ in (2.3) yields the joint probability distribution of the random variables.

The joint probability distributions for fixed \mathbf{k}_μ result from Bayes's theorem (Feller, 1957), with $P(\mathbf{k}_\mu) = 1$:

$$P(F_\mu | \mathbf{k}_\mu) = P_\infty(F_\mu) P(\mathbf{k}_\mu | F_\mu) \quad (2.5)$$

where $P_\infty(F_\mu)$ is the central limit theorem result that is independent of \mathbf{k}_μ and $P(\mathbf{r}_\nu)$. This is the first Bayesian inversion; a second similar Bayesian inversion needed to fix the intensities is introduced later. The \mathbf{k}_μ on the left-hand side are a subset of the measured reflections indicated by integers (h, k, l):

$$\mathbf{k}_\mu \subset 2\pi(h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*). \quad (2.6)$$

With the definitions

$$\sigma_n \equiv \sum_{\nu} f_{\nu}^n \quad (2.7)$$

$$E_{\mu} \equiv \sigma_2^{-1/2} F_{\mu} \quad (2.8)$$

the central limit theorem result for general-valued structure factors is (Wilson, 1949)

$$P_{\infty}(E_{\mu}) = \prod_{\mu} \{ \pi^{-1} |E_{\mu}| \exp(-|E_{\mu}|^2) \}. \quad (2.9)$$

The structure factors to be substituted in space group P_1 are

$$F(\mathbf{k}_\mu, \mathbf{r}_\nu) = \sum_{\nu} f_{\nu} \exp(i\mathbf{k}_\mu \cdot \mathbf{r}_\nu). \quad (2.10)$$

Each summand of the exponential of the integrand (2.3) can be expressed with (2.10):

$$\begin{aligned} & |F_{\mu} - F(\mathbf{k}_{\mu}, \mathbf{r}_{\nu})|^2 \\ &= |F_{\mu}|^2 + \sigma_2 + 2 \sum_{\nu > \kappa} f_{\nu} f_{\kappa} \cos[\mathbf{k}_{\mu} \cdot (\mathbf{r}_{\nu} - \mathbf{r}_{\kappa})] \\ & \quad - 2|F_{\mu}| \sum_{\nu} f_{\nu} \cos[\arg(F_{\mu}) - \mathbf{k}_{\mu} \cdot \mathbf{r}_{\nu}]. \end{aligned} \quad (2.11)$$

This result contains internal and external terms that can be associated with the partition function of a real gas in an external field (see § 4). After evaluating integral (2.3) and solving (2.4), one can derive the following expected values with $d \log [f(x)] / dx = [df(x)/dx]/f(x)$:

$$\langle |F_{\mu} - F(\mathbf{k}_{\mu}, \mathbf{r}_{\nu})|^2 \rangle = -\partial \log P(\mathbf{k}_{\mu} | F_{\mu}) / \partial \beta_{\mu}. \quad (2.12)$$

Once the β_μ have been solved by (2.4), this identity yields expected values of the radial and tangential parts of the structure factors as demonstrated in the following sections. We conclude that if the integral with the delta function (1.3) is interpreted as the micro-canonical integral of an isolated gas of point atoms with respect to the generalized complex energies $F(\mathbf{k}_\mu, \mathbf{r}_\nu)$, then the integral with the exponential (2.3) can be considered as the partition function of that gas in thermal equilibrium, where maximum entropy is obtained by determining the β_μ from the ME equations (2.4), using the maximum of the distribution as the constraint. In the following these formulas are used to derive joint probability distributions in P_1 for the single structure-factor case and for the triplet and quartet phase sums, including atomicity as well as interatomic correlations.

3. Evaluation for a single structure factor in P_1

Evaluating integral (2.3) for a single structure factor in P_1 with a uniform *a priori* distribution $P(\mathbf{r}_\nu) = 1$, the double summation term averages out. The result is (Kronenburg *et al.*, 1990).

$$P(F|\mathbf{k}) \propto P_\infty(F) \exp[-\beta(|F|^2 + \sigma_2)] \prod_\nu I_0(2\beta f_\nu |F|) \quad (3.1)$$

where I_n are the modified Bessel functions (Watson, 1952). Here the following ratio of modified Bessel functions is defined:

$$\gamma_i(x) \equiv I_i(x)/I_0(x). \quad (3.2)$$

The β is solved by (2.4):

$$|F|^{\text{obs}} = \sum_\nu f_\nu \gamma_1(2\beta f_\nu |F|^{\text{obs}}). \quad (3.3)$$

The result for maximum entropy in direct space for equal atoms is equivalent [Bricogne, 1984, equation (3.14)]; for unequal atoms it is equivalent with the multichannel solution (Bricogne, 1988, § 2.2). The ME equations (2.4) are thus confirmed in the single structure-factor case. Expanding the modified Bessel functions in (3.3) and substituting the expected value for the power of the structure-factor magnitude $\langle |E|^2 \rangle = 1$, the β for a single structure-factor results in

$$\beta \approx \left\{ \sum_\nu f_\nu^2 (1 + \sigma_2^{-1} f_\nu^2)^{-1} \right\}^{-1}. \quad (3.4)$$

Equation (3.1) with the β substituted from (3.4) is a probability distribution that can be compared with Edgeworth expansion results (Peschar & Schenk, 1987; Peschar, 1987), resulting in Figs. 1 and 2. We conclude that the maximum-entropy equations (2.4) result in a probability distribution for a single structure factor that is equivalent with Edgeworth-expansion results for small $|E|$ values. However, for larger $|E|$ values the ME distribution goes to zero in

contrast with the expansion curves, thus yielding a reliable result everywhere. For some $|E|$ values the ME result is significantly larger than the expansion result, an effect that increases with decreasing N . The exact form of the distribution is apparently not conserved by this maximum-entropy formalism; in the following sections, however, it will appear that effects of interatomic correlations present in every structure may be larger than this inconsistency. Expanding the modified Bessel functions in (3.1) and substituting (3.4), one obtains the following limit:

$$\lim_{N \rightarrow \infty} P(F) = P_\infty(F). \quad (3.5)$$

Application of (2.12) to (3.1) results in the expected value for a single structure-factor invariant:

$$\langle \cos[\arg(F^{\text{obs}}) - \mathbf{k} \cdot \mathbf{r}_\nu] \rangle = \gamma_1(2\beta f_\nu |F|^{\text{obs}}). \quad (3.6)$$

This is a completely acceptable result since $\gamma_1(x)$ is a monotonously increasing function of x smaller than one and for equal atoms $\beta f |F| \approx |E|/N^{1/2}$.

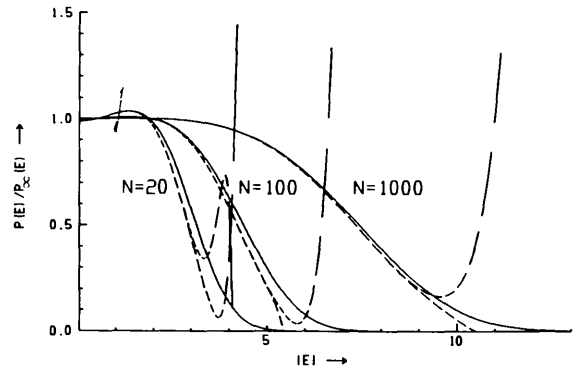


Fig. 1. Comparison of (3.1) and (3.4) with polynomial expansions (Peschar & Schenk, 1987). The solid curves represent function (3.1) for three values of N and for each one the broken curves represent the expansions in Laguerre polynomials up to $R_{9,9}$ and $R_{12,12}$.

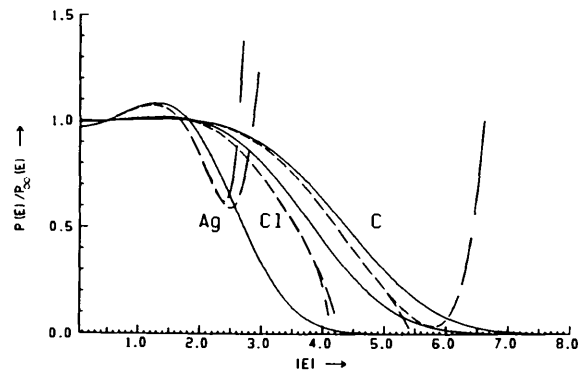


Fig. 2. Comparison of (3.1) and (3.4) with polynomial expansions as in Fig. 1, but for unequal scattering factors f_ν . 95 carbon atoms and 5 atoms of varying elements (see figure) are present.

4. Interatomic correlations and cluster integrals in $P1$

Having evaluated integral (2.3) for the single structure-factor case with a uniform *a priori* distribution $P(\mathbf{r}_\nu) = 1$, we demonstrate how integral (2.3) can be calculated with a non-uniform *a priori* atomic distribution that accounts for interatomic distances as well as atomicity. A set of minimal interatomic distances $D_{\nu\kappa}$ for each pair of atoms is introduced and easily extended to interatomic distances $d_{\nu\alpha}$ for each pair of bonded atoms later. Derivation of radial correlation between atoms starts with the hard-sphere *a priori* probability distribution:

$$\begin{cases} P(\mathbf{r}_\nu) = 1 & \text{when } |\mathbf{r}_\nu - \mathbf{r}_\kappa| > D_{\nu\kappa} \text{ for each } \nu \\ & \text{and each } \kappa \neq \nu \\ P(\mathbf{r}_\nu) = 0 & \text{otherwise.} \end{cases} \quad (4.1)$$

In order to apply the irreducible Mayer cluster integrals (Mayer & Mayer, 1940), the following definitions are used in correspondence with (2.11); the internal terms in integral (2.3) are

$$\begin{cases} \Omega(\mathbf{r}_\nu - \mathbf{r}_\kappa) \equiv \exp \left\{ -2f_\nu f_\kappa \sum_\mu \beta_\mu \cos [\mathbf{k}_\mu \cdot (\mathbf{r}_\nu - \mathbf{r}_\kappa)] \right\} \\ & \text{when } |\mathbf{r}_\nu - \mathbf{r}_\kappa| > D_{\nu\kappa} \\ \Omega(\mathbf{r}_\nu - \mathbf{r}_\kappa) \equiv 0 & \text{otherwise.} \end{cases} \quad (4.2)$$

The external terms in integral (2.3) are

$$\omega(\mathbf{r}_\nu) \equiv \exp \left\{ 2f_\nu \sum_\mu \beta_\mu |F_\mu| \cos [\arg(F_\mu) - \mathbf{k}_\mu \cdot \mathbf{r}_\nu] \right\}. \quad (4.3)$$

The evaluation of integral (2.3) with (2.11) substituted for $P1$ is equivalent to the evaluation of formula (11) in the paper by Van Kampen (1961), in which for the internal potential φ the double-summation terms $\Omega(\mathbf{r}_\nu - \mathbf{r}_\kappa)$ are substituted, supplied by a hard-sphere potential as in definition (4.2), and for the external potential ψ the single summation terms $\omega(\mathbf{r}_\nu)$ as defined in (4.3). Evaluation of integral (2.3) is equivalent to the evaluation of the partition function for a classical hard-sphere gas in an external field, which can be done by means of irreducible Mayer cluster integrals (Van Kampen, 1961). The maximum-entropy equations (2.4) are again solved for this new result of integral (2.3).

Application is only exactly correct in the thermodynamic limit ($N \rightarrow \infty$, $N/V = n$). Since usage of integral (2.3) is already consistent for relatively small N (see Fig. 1) in the uniform case, it is assumed that $N/V = n$ is small enough in practical cases to ensure that the irreducible cluster integral results hold.

The first and second irreducible Mayer cluster integrals as used by Van Kampen in order to evaluate

integral (2.3) are:

$$\varepsilon_{1\nu\kappa} = \int \int_{\mathbf{r}_\nu \mathbf{r}_\kappa} \omega(\mathbf{r}_\nu) \omega(\mathbf{r}_\kappa) \{ \Omega(\mathbf{r}_\nu - \mathbf{r}_\kappa) - 1 \} d\mathbf{r}_\nu d\mathbf{r}_\kappa \quad (4.4)$$

$$\begin{aligned} \varepsilon_{2\nu\kappa\lambda} = & \int \int \int_{\mathbf{r}_\nu \mathbf{r}_\kappa \mathbf{r}_\lambda} \omega(\mathbf{r}_\nu) \omega(\mathbf{r}_\kappa) \omega(\mathbf{r}_\lambda) \\ & \times \{ \Omega(\mathbf{r}_\nu - \mathbf{r}_\kappa) - 1 \} \{ \Omega(\mathbf{r}_\kappa - \mathbf{r}_\lambda) - 1 \} \\ & \times \{ \Omega(\mathbf{r}_\nu - \mathbf{r}_\lambda) - 1 \} d\mathbf{r}_\nu d\mathbf{r}_\kappa d\mathbf{r}_\lambda. \end{aligned} \quad (4.5)$$

With these definitions, for equal atoms, incorporation of (4.1) in integral (2.3) results according to Van Kampen's formula (Van Kampen, 1961, equation 14) in

$$\begin{aligned} P(\mathbf{k}_\mu | F_\mu) = & \prod_\mu \left\{ \exp(-\beta_\mu |F_\mu|^2) \prod_\nu I_0(2\beta_\nu f_\nu |F_\mu|) \right\} \\ & \times \exp \left[\binom{N}{2} (V\langle\omega\rangle)^{-2} \varepsilon_1 \right. \\ & \left. + \binom{N}{3} (V\langle\omega\rangle)^{-3} \varepsilon_2 + \dots \right]. \end{aligned} \quad (4.6)$$

In the unequal-atoms case, the $\binom{N}{2}$ in this formula results from double summation of the $\varepsilon_{1\nu\kappa}$ over ν and κ and the $\binom{N}{3}$ from threefold summation *etc.* Higher-order cluster integrals may be defined analogously and substituted in (4.6). Evaluation of ε_1 and ε_2 for a single structure factor in $P1$ yields, taking $f_\lambda \approx f_\kappa$ for the second cluster integral (see Appendix 1)

$$\langle\omega\rangle \approx \langle\omega_\nu\rangle = I_0(2\beta f_\nu |F|) \quad (4.7)$$

$$\begin{aligned} G_n(k, D_{\nu\kappa}) \equiv & 4\pi D_{\nu\kappa} (nk)^{-2} [\sin(nkD_{\nu\kappa})/nkD_{\nu\kappa} \\ & - \cos(nkD_{\nu\kappa})] \end{aligned} \quad (4.8)$$

$$\begin{aligned} \varepsilon_{1\nu\kappa} \approx & 2V\langle\omega\rangle^2 \{ I_1(2\beta f_\nu f_\kappa) - I_0(2\beta f_\nu f_\kappa) \\ & \times \gamma_1(2\beta f_\nu |F|) \gamma_1(2\beta f_\kappa |F|) \} G_1(k, D_{\nu\kappa}) \end{aligned} \quad (4.9)$$

$$\varepsilon_{2\nu\kappa\lambda} \approx -4V\langle\omega\rangle^3 \beta f_\nu f_\kappa \{ I_0(2\beta f_\nu f_\kappa) G_1(k, D_{\nu\kappa}) \}^2. \quad (4.10)$$

Substituted in (4.6), these results change the single structure-factor distribution (3.1) as well as the expected value (3.6). G_1 can be interpreted as the excluded volume by the minimal interatomic distance as seen from \mathbf{k} , which can be demonstrated by taking a limit

$$\lim_{k \rightarrow 0} G_1(k, D) = \frac{4}{3}\pi D^3. \quad (4.11)$$

The ε_1 and ε_2 change sign as the corresponding minimal interatomic distances make positive or negative contributions to the distribution in reciprocal space on the sphere $|\mathbf{k}| = k$. After substitution of ε_1 and ε_2 in (4.6), the expected value becomes with

(2.12) and $I_0 = 1$, $I_2 = 0$:

$$\left\{ \begin{aligned} & \langle |F(\mathbf{k}, \mathbf{r}_\nu)|^2 \rangle \\ & = \sum_\nu f_\nu \left\{ f_\nu - V^{-1} \sum_\kappa f_\kappa \right. \\ & \quad \times \left. \{ 1 - [2(N-2)/3V] G_1(k, D_{\nu\kappa}) \} G_1(k, D_{\nu\kappa}) \right\} \\ & \langle \cos [\arg (F^{\text{obs}}) - \mathbf{k} \cdot \mathbf{r}_\nu] \rangle \\ & = \gamma_1(2\beta f_\nu |F|^{\text{obs}}) \left\{ 1 - (V f_\nu)^{-1} \sum_\kappa f_\kappa G_1(k, D_{\nu\kappa}) \right\}. \end{aligned} \right. \quad (4.12)$$

Once the irreducible cluster integral for a sphere has been computed, the result for a spherical layer is easily obtained. If a number $n_{\nu\alpha}$ of atoms lies within a spherical layer of a small thickness τ with radius $d_{\nu\alpha}$ from atom ν , it can be demonstrated that the Debye scattering equation (Debye, 1915) follows from the first irreducible cluster integral. The τ must be given by conservation of probability:

$$4\pi d_{\nu\alpha}^2 \tau V^{-1} = n_{\nu\alpha}. \quad (4.13)$$

The $n_{\nu\alpha}$ are defined in such a way that

$$\left\{ \begin{aligned} & n_{\nu\nu} = 1 \quad \text{for all } \nu \in 1 \dots N \text{ and} \\ & \sum_\alpha n_{\nu\alpha} = N - m_\nu \end{aligned} \right. \quad (4.14)$$

where m_ν is the number of atoms that do not bind to atom ν . These atoms must be at some distance from atom ν larger than some minimal interatomic distance $D_{\nu\kappa}$. The interatomic distances $d_{\nu\alpha}$ are not specified further except for

$$d_{\nu\nu} = 0 \quad \text{for all } \nu \in 1 \dots N. \quad (4.15)$$

The contribution of the spherical layer to the cluster integrals is given by

$$-\tau \, dG_n(k, D)/dD = \tau 4\pi d_{\nu\alpha}^2 \sin(nkd_{\nu\alpha})/nkd_{\nu\alpha}. \quad (4.16)$$

Substitution of (4.13) and (4.16) in (4.12) gives

$$\left\{ \begin{aligned} & \langle |F(\mathbf{k}, \mathbf{r}_\nu)|^2 \rangle = \sum_\nu f_\nu \left\{ f_\nu + \sum_{\alpha \neq \nu} n_{\nu\alpha} f_\alpha \sin(kd_{\nu\alpha})/kd_{\nu\alpha} \right. \\ & \quad - V^{-1} \sum_{\kappa=1}^{m_\nu} \left. \{ 1 - [2(N-2)/3V] \right. \\ & \quad \times \left. G_1(k, D_{\nu\kappa}) \} f_\kappa G_1(k, D_{\nu\kappa}) \right\} \\ & \langle \cos [\arg (F^{\text{obs}}) - \mathbf{k} \cdot \mathbf{r}_\nu] \rangle \\ & = \gamma_1(2\beta f_\nu |F|^{\text{obs}}) \\ & \quad \times \left\{ 1 + f_\nu^{-1} \left[\sum_{\alpha \neq \nu} n_{\nu\alpha} f_\alpha \sin(kd_{\nu\alpha})/kd_{\nu\alpha} \right. \right. \\ & \quad \left. \left. - V^{-1} \sum_{\kappa=1}^{m_\nu} f_\kappa G_1(k, D_{\nu\kappa}) \right] \right\}. \end{aligned} \right. \quad (4.17)$$

If one takes only the terms dependent on $|F|$ in (4.6), the corresponding j.p.d. is

$$\begin{aligned} P(F|\mathbf{k}) & \propto P_\infty(F) \exp(-\beta|F|^2) \\ & \times \prod_\nu \left(I_0(2\beta f_\nu |F|) \exp \left[\gamma_1(2\beta f_\nu |F|) \right. \right. \\ & \times \left. \left. \left\{ \sum_{\alpha \neq \nu} n_{\nu\alpha} [\sin(kd_{\nu\alpha})/kd_{\nu\alpha}] \gamma_1(2\beta f_\alpha |F|) \right. \right. \right. \\ & \left. \left. \left. - V^{-1} \sum_{\kappa=1}^{m_\nu} G_1(k, D_{\nu\kappa}) \gamma_1(2\beta f_\kappa |F|) \right\} \right] \right). \end{aligned} \quad (4.18)$$

The Debye (1915) scattering equation for interatomic distances between binding atoms is present in the first identity of (4.17) as expected, supplied by a term that accounts for minimal interatomic distances of pairs of atoms that do not bind. This accounts well for the effect of minimal interatomic distances and interatomic distances on the Wilson statistics; the *a priori* knowledge (4.1) might correspond well with the actual structure present in the unit cell and averaging over $|\mathbf{k}| = k$ yields expected values as functions of k . The effect of interatomic distances on the Wilson curves has been the subject of studies (Subramanian & Hall, 1982) using the Debye scattering equation for all interatomic distances present in the structure. The additional term in (4.17) for the minimal interatomic distances, for example between atoms from two different fragments, might well account for the differences of the Debye scattering equation with experimental Wilson curves; the values to be substituted for the $D_{\nu\kappa}$ remain however to be found. In Fig. 3 the first identity in (4.17) is compared with randomly generated structures with a minimal interatomic distance (no Debye contribution). The minimal interatomic distance used for the random

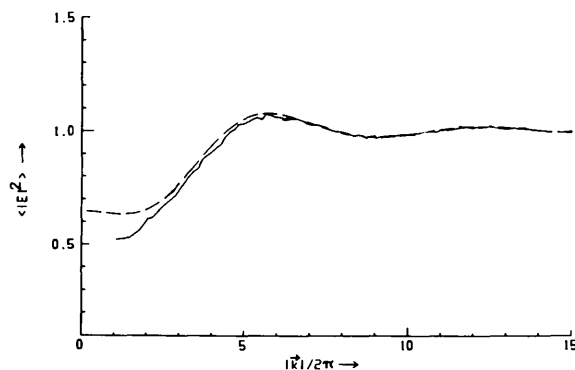


Fig. 3. Comparison of the expected value $\langle |E|^2 \rangle$ as a function of the length of the reciprocal vector $|\mathbf{k}|$ with equal $f_\nu = f$ for $N = 50$ and $V = 18 N \text{ \AA}^3$. A hard-sphere model with $D = 1.54 \text{ \AA}$ is assumed (no Debye contribution), leading to the function (4.17) with $n_{\nu\alpha} = 0$ (broken curve). The solid curve results from 10 000 random structures with $D = 1.54 \text{ \AA}$.

structures is consistent with the theoretical curve (4.17) for the corresponding D ; only for very small $|\mathbf{k}|$ values do irreducible cluster integrals of order larger than two give a small contribution. Although the β does not appear in the first identity of (4.17), for other results the β should be solved from (2.4) and (4.18), resulting in (with $I_0 \approx 1$, $I_2 = 0$)

$$|F|^{\text{obs}} = \sum_{\nu} \gamma_1(2\beta f_{\nu}|F|^{\text{obs}}) \times \left\{ f_{\nu} + \sum_{\alpha \neq \nu} n_{\nu\alpha} f_{\alpha} \sin(kd_{\nu\alpha})/kd_{\nu\alpha} - V^{-1} \sum_{\kappa=1}^{m_{\nu}} f_{\kappa} G_1(k, D_{\nu\kappa}) \right\}. \quad (4.19)$$

As is demonstrated in the following sections, this equation is important for the effect of the fluctuations in the Wilson curve on the $|E|^{\text{obs}}$ values to be substituted in the exponential multiplet distributions.

5. The triplet distribution from ME and the cluster integral

In the following it is demonstrated that evaluation of integral (2.3) by means of irreducible cluster integrals and application of ME equations (2.4) leads to exponential joint probability distributions for multiplets. These distributions are consistent with maximum entropy and interatomic correlations and result directly from the first irreducible cluster integral. The evaluation of the first irreducible cluster integral (4.4) for three structure factors $\{F_{\mu}, \mu = 1, 2, 3\}$ in $P1$ yields extra triplet terms [$s \in (1, 2)$ with $f_1 = f_{\nu}$ and $f_2 = f_{\kappa}$, $n_{\mu s}$ six indices and $\mathbf{r} = \mathbf{r}_{\nu} - \mathbf{r}_{\kappa}$]:

$$\langle \omega \rangle = \langle \omega_{\nu} \rangle = \prod_{\mu} I_0(2\beta_{\mu} f_{\nu} |F_{\mu}|) \quad (5.1)$$

$$\begin{aligned} \varepsilon_{1\nu\kappa} = & -\langle \omega \rangle^2 \left\{ \prod_{\mu} I_0(2\beta_{\mu} f_{\nu} f_{\kappa}) \right\} \\ & \times \sum_{n_{\mu s} = -\infty}^{\infty} \left\{ \prod_{\mu s} \gamma_{n_{\mu s}}(2\beta_{\mu} f_s |F_{\mu}|) \right\} \\ & \times \int_{\mathbf{r}_1} \int_{|\mathbf{r}| < D_{\nu\kappa}} \exp \left(i \left\{ \sum_{\mu s} n_{\mu s} \arg(F_{\mu}) \right. \right. \\ & \left. \left. - \sum_{\mu} [n_{\mu 1} \mathbf{k}_{\mu} \cdot \mathbf{r}_1 + n_{\mu 2} \mathbf{k}_{\mu} \cdot (\mathbf{r}_1 - \mathbf{r})] \right\} \right) d\mathbf{r} d\mathbf{r}_1. \quad (5.2) \end{aligned}$$

Evaluation of this integral with $I_0(2\beta f^2) \approx 1$ and $\gamma_2(x) \approx 0$ yields the sum of the three single structure-factor cluster-integral results (4.9), plus an extra trip-

let term when $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{0}$:

$$\begin{aligned} \varepsilon_{1\nu\kappa}^{\text{trip}} = & -2V\langle \omega \rangle^2 \cos(\psi_t) \\ & \times \left\{ \frac{4}{3}\pi D_{\nu\kappa}^3 + \sum_{\mu} G_1(k_{\mu}, D_{\nu\kappa}) \right\} \\ & \times \left\{ \prod_{\mu} \gamma_1(2\beta_{\mu} f_{\nu} |F_{\mu}|) + \prod_{\mu} \gamma_1(2\beta_{\mu} f_{\kappa} |F_{\mu}|) \right\} \quad (5.3) \end{aligned}$$

with $\psi_t \equiv \arg(F_1) + \arg(F_2) + \arg(F_3)$. The twofold summation in (4.6) is simplified to a single summation:

$$\begin{aligned} \sum_{\nu, \kappa} \left\{ \prod_{\mu} \gamma_1(2\beta_{\mu} f_{\nu} |F_{\mu}|) + \prod_{\mu} \gamma_1(2\beta_{\mu} f_{\kappa} |F_{\mu}|) \right\} \\ = (N-1) \sum_{\nu, \mu} \prod_{\nu, \mu} \gamma_1(2\beta_{\mu} f_{\nu} |F_{\mu}|). \quad (5.4) \end{aligned}$$

A sum of cluster terms leads with (4.6) to a product of probabilities.

Now the following identity is used:

$$P(\psi_t, |F_{\mu}| | \mathbf{k}_{\mu}) = P(|F_{\mu}| | \mathbf{k}_{\mu}) P(\psi_t | |F_{\mu}|, \mathbf{k}_{\mu}). \quad (5.5)$$

This is the second Bayesian inversion, needed to fix the intensities. As the total irreducible cluster integral of the triplet is the sum of the three single structure-factor irreducible cluster integrals and a term dependent on the triplet phase sum ψ_t , only this triplet term appears in $P(\psi_t | |F_{\mu}|, \mathbf{k}_{\mu})$ which is the conditional probability distribution. Application of (4.13), (4.16) and (4.11) to all terms in (5.3) results in

$$\begin{aligned} P(\psi_t | |F_{\mu}|, \mathbf{k}_{\mu}) \\ \propto \exp \left[2 \cos(\psi_t) \sum_{\nu} \left(1 + f_{\nu}^{-1} \right. \right. \\ \left. \left. \times \sum_{\mu} \left\{ \sum_{\alpha \neq \nu} n_{\nu\alpha} f_{\alpha} \sin(k_{\mu} d_{\nu\alpha}) / k_{\mu} d_{\nu\alpha} \right. \right. \right. \\ \left. \left. \left. - V^{-1} \sum_{\kappa=1}^{m_{\nu}} f_{\kappa} G_1(k_{\mu}, D_{\nu\kappa}) \right\} \right) \prod_{\mu} \gamma_1(2\beta_{\mu} f_{\nu} |F_{\mu}|) \right]. \quad (5.6) \end{aligned}$$

The set of three equations (2.4) are now dependent on ψ_t , so the β_{μ} need to be solved as well. The β_{μ} are known as functions of ψ_t and $|F_{\mu}|^{\text{obs}}$ and are averaged over ψ_t , in contrast with the single structure factor β which was averaged over $|F|$. Substitution in (5.6) yields the final triplet distribution. However, when the terms in (2.4) that are dependent on ψ_t are neglected, (2.4) results in (4.19) for each μ . When these are substituted in (5.6), the terms with summation over μ cancel out. From this it can be concluded that the interatomic correlations responsible for the fluctuation in the Wilson curve [first identity of (4.17)] do not affect the resulting triplet main exponential term when maximum entropy is assumed. This means that renormalization of the $|E|^{\text{obs}}$ values for each

reciprocal spherical layer, which is an option in many phasing programs, is in contradiction with these maximum-entropy results when using triplet phasing; the original $|E|^{\text{obs}}$ values normalized with respect to the total measured reciprocal space should be used. This results in

$$P(\psi_r || F_\mu, \mathbf{k}_\mu) \propto \exp \left[2 \cos(\psi_r) \sum_{\nu} \prod_{\mu} \gamma_1(2\beta_\mu f_\nu | F_\mu) \right]. \quad (5.7)$$

Using $|E|$ definition (2.8), approximating (3.4) by $\beta_\mu \approx \sigma_2^{-1}$ and using $\gamma_1(2x) \approx x$, substitution in (5.7) yields the Cochran (1955) result:

$$P(\psi_r || F_\mu^{\text{obs}}, \mathbf{k}_\mu) \propto \exp \left[2 \cos(\psi_r) \sigma_3 \sigma_2^{-3/2} \prod_{\mu} |E_\mu|^{\text{obs}} \right]. \quad (5.8)$$

It is concluded that the maximum-entropy method with ME equations (2.4) is consistent with the well known exponential Cochran result. However, the approximations made also indicate that the full exponential result is more complicated. As the expansion of the γ_1 functions in (5.7) contains only powers larger than 0, an $|F_1 F_2 F_3|^{\text{obs}}$ factor is present in every power, which is in accordance with the exponential Edgeworth triplet result (Karle & Gilardi, 1973).

6. Higher-order exponential triplet terms and the triplet inequality

In order to derive the higher-order exponential triplet terms, the β_μ have to be solved from the three interdependent ME equations (2.4), averaging now over the triplet-invariant random variable ψ_r , in contrast with averaging over the $|E|$ value as was the case for a single structure factor. Expansion of the modified Bessel functions may yield the correct exponential terms. For simplicity U values are defined as

$$U_\mu \equiv \sigma_2^{-1} |F_\mu|^{\text{obs}} = \sigma_2^{-1/2} |E_\mu|^{\text{obs}}. \quad (6.1)$$

Evaluating in the triplet case the set of three ME equations (2.4) and substituting also the ψ_r -dependent part (5.7) (taking $D \rightarrow 0$: no correlations for $\nu \neq \nu'$), one obtains the following set of equations [using $\gamma_1(2x) \approx x(1 - x^2/2)$]:

$$\begin{cases} \sigma_2^{-1} U_1 = \beta_1 U_1 - \frac{1}{2} \sigma_2 \sigma_4 \beta_1^3 U_1^3 \\ \quad + \beta_2 \beta_3 U_2 U_3 \sigma_3 \cos(\psi_r) \\ \sigma_2^{-1} U_2 = \beta_2 U_2 - \frac{1}{2} \sigma_2 \sigma_4 \beta_2^3 U_2^3 \\ \quad + \beta_1 \beta_3 U_1 U_3 \sigma_3 \cos(\psi_r) \\ \sigma_2^{-1} U_3 = \beta_3 U_3 - \frac{1}{2} \sigma_2 \sigma_4 \beta_3^3 U_3^3 \\ \quad + \beta_1 \beta_2 U_1 U_2 \sigma_3 \cos(\psi_r). \end{cases} \quad (6.2)$$

In order to solve this set of equations, the ψ_r -dependent terms are considered to be an offset vector. Identifying this offset vector with the first-order differential matrix in the Taylor expansion of the three functions in the β_μ variables, the set of equations may be solved by inverting the differential matrix:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \approx \begin{pmatrix} \beta_1' \\ \beta_2' \\ \beta_3' \end{pmatrix} - \begin{pmatrix} 1 & \sigma_3 \beta_3 \frac{U_2 U_3}{U_1} \cos(\psi_r) & \sigma_3 \beta_2 \frac{U_2 U_3}{U_1} \cos(\psi_r) \\ \sigma_3 \beta_3 \frac{U_1 U_3}{U_2} \cos(\psi_r) & 1 & \sigma_3 \beta_1 \frac{U_1 U_3}{U_2} \cos(\psi_r) \\ \sigma_3 \beta_2 \frac{U_1 U_2}{U_3} \cos(\psi_r) & \sigma_3 \beta_1 \frac{U_1 U_2}{U_3} \cos(\psi_r) & 1 \end{pmatrix}^{-1} \times \begin{pmatrix} \beta_2 \beta_3 (U_2 U_3 / U_1) \\ \beta_1 \beta_3 (U_1 U_3 / U_2) \\ \beta_1 \beta_2 (U_1 U_2 / U_3) \end{pmatrix} \sigma_3 \cos(\psi_r). \quad (6.3)$$

The β'_μ are defined as (using now $\beta_\mu \approx \sigma_2^{-1}$)

$$\beta'_\mu = \sigma_2^{-1} (1 + \frac{1}{2} \sigma_2^{-1} \sigma_4 U_\mu^2). \quad (6.4)$$

The determinant of the matrix cannot be zero when this solution is not singular. As the determinant is positive for $U_1 = U_2 = U_3 \rightarrow 0$, it must be positive for every valid (U_1, U_2, U_3, ψ_r) combination:

$$\|\nabla\| = 1 - \sigma_2^{-2} \sigma_3^2 \cos^2(\psi_r) \{U_1^2 + U_2^2 + U_3^2 - 2\sigma_2^{-1} \sigma_3 U_1 U_2 U_3 \cos(\psi_r)\} > 0. \quad (6.5)$$

Because $\cos^2(\psi_r) \leq 1$, this is always true in the case of equal and positive f_ν when the Karle-Hauptman inequality (Karle & Hauptman, 1950) for three structure factors is fulfilled:

$$2U_1 U_2 U_3 \cos(\psi_r) > U_1^2 + U_2^2 + U_3^2 - 1. \quad (6.6)$$

It can be concluded that this maximum-entropy solution leads to a weaker probabilistic inequality that may be compared with the Karle-Hauptman triplet inequality. Apparently, for the values at the maximum of the distribution as defined by (2.4), the inequalities are weaker than but close to those derived for expected values (Bricogne, 1984, § 6.1.2). For derivation of the exponential multiplet distributions the following matrix inversion formula is used:

$$(I + A)^{-1} = I - A + A^2 - A^3 + \dots \quad (6.7)$$

Although this power series need not be convergent for every ψ_r , it can be averaged over ψ_r using

$$\langle \cos^{2n}(\psi) \rangle = 2^{-2n} \binom{2n}{n}. \quad (6.8)$$

The following β_μ values result (up to order N^{-2}):

$$\begin{aligned} \beta_1 &= \sigma_2^{-1} \{1 + \frac{1}{2} \sigma_2^{-1} \sigma_4 U_1^2 + \frac{1}{2} \sigma_2^{-2} \sigma_3^2 (U_2^2 + U_3^2)\} \\ \beta_2 &= \sigma_2^{-1} \{1 + \frac{1}{2} \sigma_2^{-1} \sigma_4 U_2^2 + \frac{1}{2} \sigma_2^{-2} \sigma_3^2 (U_1^2 + U_3^2)\} \\ \beta_3 &= \sigma_2^{-1} \{1 + \frac{1}{2} \sigma_2^{-1} \sigma_4 U_3^2 + \frac{1}{2} \sigma_2^{-2} \sigma_3^2 (U_1^2 + U_2^2)\}. \end{aligned} \quad (6.9)$$

Finally, the β_μ have to be corrected for the fact that $P(|F_\mu|) = P(|F_1|)P(|F_2|)P(|F_3|)$ no longer. Integration of (5.8) over ψ_i gives

$$\begin{aligned} P(|F_\mu| | \mathbf{k}_\mu) &\propto P(|F_1| | \mathbf{k}_1) P(|F_2| | \mathbf{k}_2) P(|F_3| | \mathbf{k}_3) \\ &\quad \times I_0(2\sigma_3\beta_1\beta_2\beta_3|F_1F_2F_3|). \end{aligned} \quad (6.10)$$

Using $\log [I_0(2x)] \approx x^2$ and $\langle |E_\mu|^2 \rangle = 1$ since no phases are involved here, from all the β_μ , $\sigma_2^{-4}\sigma_3^2$ has to be subtracted. Substitution in (5.7) and using again $\gamma_1(2x) \approx x(1 - x^2/2)$:

$$\begin{aligned} P(\psi_i | |F_\mu|^{\text{obs}}, \mathbf{k}_\mu) &\propto \exp(2 \cos(\psi_i) U_1 U_2 U_3 \\ &\quad \times \{\sigma_3 - 3\sigma_2^{-3}\sigma_3^3 + [\sigma_2^{-2}\sigma_3^3 + \frac{1}{2}(\sigma_2^{-1}\sigma_3\sigma_4 - \sigma_5)] \\ &\quad \times (U_1^2 + U_2^2 + U_3^2)\}). \end{aligned} \quad (6.11)$$

Comparing this with the exponential Edgeworth triplet result [Karle & Gilardi, 1973, equation (1)], it can be concluded that most $N^{-3/2}$ terms are present. The extra terms in the Edgeworth result may result from the fact that the third term in (2.11) has been neglected in the derivation of the triplet cluster integral result (5.2).

7. Exponential quartet terms

Starting with a set of ME equations of an arbitrary number of intensities (2.4), a summation over all invariants of irreducible cluster integrals like (5.2) results. For the quartet phase sum the 7×7 differential matrix of the set of seven equations (2.4) can easily be computed and inversion and substitution again result in an exponential distribution. The general exponential expression is, taking no interatomic correlations into account, with i the index over the invariant multiplets ψ_i and μ_i the index over the main intensities corresponding with ψ_i :

$$\begin{aligned} P(\psi_i | |F_\mu|, \mathbf{k}_\mu) &\propto \exp \left[2 \sum_i \cos(\psi_i) \sum_{\nu, \mu_i} \prod \gamma_1(2\beta_{\mu_i} f_\nu | F_{\mu_i}) \right]. \end{aligned} \quad (7.1)$$

In the quartet case seven terms are taken into account in (7.1) corresponding with one quartet phase sum and six triplets. The resulting offset vector is [denoting indices 1, ..., 7 for $\mathbf{h}, \mathbf{k}, \mathbf{l}, -\mathbf{h}-\mathbf{k}-\mathbf{l}, \mathbf{h}+\mathbf{k}, \mathbf{h}+\mathbf{l}, \mathbf{k}+\mathbf{l}$

respectively, ψ_q for the quartet phase sum and ψ_1, \dots, ψ_6 for the six corresponding triplet phase sums (125), (345), (136), (246), (147), (237) respectively, using $\gamma_1(2x) \approx x$ and $\beta_\mu \approx \sigma_2^{-1}$]:

$$\begin{aligned} &\left[\begin{aligned} &U_1^{-1} \{ \sigma_2^{-3} \sigma_4 U_2 U_3 U_4 \cos(\psi_q) + \sigma_2^{-2} \sigma_3 [U_2 U_5 \cos(\psi_1) \\ &\quad + U_3 U_6 \cos(\psi_3) + U_4 U_7 \cos(\psi_5)] \} \\ &U_2^{-1} \{ \sigma_2^{-3} \sigma_4 U_1 U_3 U_4 \cos(\psi_q) + \sigma_2^{-2} \sigma_3 [U_1 U_5 \cos(\psi_1) \\ &\quad + U_4 U_6 \cos(\psi_4) + U_3 U_7 \cos(\psi_6)] \} \\ &U_3^{-1} \{ \sigma_2^{-3} \sigma_4 U_1 U_2 U_4 \cos(\psi_q) + \sigma_2^{-2} \sigma_3 [U_4 U_5 \cos(\psi_2) \\ &\quad + U_1 U_6 \cos(\psi_3) + U_2 U_7 \cos(\psi_6)] \} \\ &U_4^{-1} \{ \sigma_2^{-3} \sigma_4 U_1 U_2 U_3 \cos(\psi_q) + \sigma_2^{-2} \sigma_3 [U_3 U_5 \cos(\psi_2) \\ &\quad + U_2 U_6 \cos(\psi_4) + U_1 U_7 \cos(\psi_5)] \} \\ &U_5^{-1} \sigma_2^{-2} \sigma_3 \{ U_1 U_2 \cos(\psi_1) + U_3 U_4 \cos(\psi_2) \} \\ &U_6^{-1} \sigma_2^{-2} \sigma_3 \{ U_1 U_3 \cos(\psi_3) + U_2 U_4 \cos(\psi_4) \} \\ &U_7^{-1} \sigma_2^{-2} \sigma_3 \{ U_1 U_4 \cos(\psi_5) + U_2 U_3 \cos(\psi_6) \} \end{aligned} \right]. \end{aligned} \quad (7.2)$$

Inversion of the 7×7 differential matrix leads to an exponential distribution as in the triplet case. The only terms of order N^{-1} however result from the lower right 3×3 diagonal submatrix, yielding the following set of β_μ values:

$$\begin{aligned} &\left\{ \begin{aligned} \beta_1 &= \sigma_2^{-1} \\ \beta_2 &= \sigma_2^{-1} \\ \beta_3 &= \sigma_2^{-1} \\ \beta_4 &= \sigma_2^{-1} \\ \beta_5 &= \sigma_2^{-1} \{ 1 - U_5^{-1} \sigma_2^{-1} \sigma_3 [U_1 U_2 \cos(\psi_1) \\ &\quad + U_3 U_4 \cos(\psi_2)] \} \\ \beta_6 &= \sigma_2^{-1} \{ 1 - U_6^{-1} \sigma_2^{-1} \sigma_3 [U_1 U_3 \cos(\psi_3) \\ &\quad + U_2 U_4 \cos(\psi_4)] \} \\ \beta_7 &= \sigma_2^{-1} \{ 1 - U_7^{-1} \sigma_2^{-1} \sigma_3 [U_1 U_4 \cos(\psi_5) \\ &\quad + U_2 U_3 \cos(\psi_6)] \}. \end{aligned} \right. \end{aligned} \quad (7.3)$$

As a $\cos(\psi_2)$ term appears in the solution of β_5 , a product $\cos(\psi_2) \cos(\psi_1)$ appears from the first triplet product in (7.1), which can be transposed in a $\cos(\psi_4)$ term using $\psi_1 + \psi_2 = \psi_4$:

$$\begin{aligned} \cos(\psi_2) \cos(\psi_1) &= \cos^2(\psi_2) \cos(\psi_4) \\ &\quad + \cos(\psi_2) \sin(\psi_2) \sin(\psi_4). \end{aligned} \quad (7.4)$$

With an average now over ψ_2 , by (6.8) the first term results in a pure $\cos(\psi_4)$ term and the second term vanishes, which confirms the cross-term principle. Similar formulae hold for the other pairs of triplets that sum up to the quartet and this is easily generalized

to higher powers of $\cos(\psi_i)$. The resulting exponential distribution up to order N^{-1} is equal to the well known exponential quartet distribution (Hauptman, 1975a, 1976):

$$P(\psi_q, \psi_1, \dots, \psi_6 | |F_\mu|^{\text{obs}}, \mathbf{k}_\mu) \\ \propto \exp \{ 2 \cos(\psi_q) U_1 U_2 U_3 U_4 (\sigma_4 - 3\sigma_2^{-1} \sigma_3^2) \\ + 2\sigma_3 [U_1 U_2 U_5 \cos(\psi_1) + \dots + U_2 U_3 U_7 \cos(\psi_6)] \}. \quad (7.5)$$

This is also the exponential Edgeworth result (Peschar, 1987). Finally, $P(\psi_q | |F_\mu|^{\text{obs}}, \mathbf{k}_\mu)$ can be obtained by integrating out the ψ_i under the restriction $\psi_1 + \psi_2 = \psi_3 + \psi_4 = \psi_5 + \psi_6 = \psi_q$ [Hauptman, 1975b, equation (2.5)]. From this it may be concluded that the well known exponential quartet distribution is consistent with maximum-entropy calculations. Higher-order exponential quartet terms can be derived in analogy with the higher-order triplet terms, taking into account the cross-term principle as mentioned.

Derivation of the quartet irreducible cluster integral is similar to the triplet case (5.2). However, after substituting the maximum-entropy solution (4.19) for each μ assuming also interatomic correlations, these cluster terms do not cancel out as in the triplet case because of the cross-term principle. The extra terms K_μ contain the remaining cluster terms:

$$K_\mu = \sigma_2^{-1} \sigma_3 \sum_\nu f_\nu^2 \left[\sum_{\alpha \neq \nu} n_{\nu\alpha} f_\alpha \sin(k_\mu d_{\nu\alpha}) / k_\mu d_{\nu\alpha} \right. \\ \left. - V^{-1} \sum_{\kappa=1}^{m_\nu} f_\kappa G_1(k_\mu, D_{\nu\kappa}) \right]. \quad (7.6)$$

These terms may be identified by the first identity of (4.17) with expected $\langle |E|^2 \rangle$ values:

$$K_\mu \approx \sigma_2^{-1} \sigma_3^2 (\langle |E_\mu|^2 \rangle - 1) \quad (7.7)$$

where the $\langle |E_\mu|^2 \rangle$ are the expected values in the reciprocal spherical layers obtained from the Wilson curve (corrected for the Debye-Waller factor).

In Appendix B it is demonstrated that, for unequal f_ν , the differences between the f_ν being relatively small, and keeping the average of the f_ν constant:

$$(\sigma_2 \sigma_4 - \sigma_3^2) / \sigma_2 \sigma_4 \approx \langle |E|^2 \rangle - 1. \quad (7.8)$$

When the substituted $|E|^{\text{obs}}$ values are normalized with respect to the total measured reciprocal space, for each individual reciprocal spherical layer the $\langle |E|^2 \rangle$ can be unequal to 1, yielding the terms (7.7). Assuming interatomic correlations for that spherical layer is equivalent to assuming unequal f_ν but keeping the average of the f_ν constant. This effect of interatomic correlations affects the exponential quartet main term. Distribution (7.5) may be corrected for radial interatomic correlations by replacing the main

quartet term with

$$\sigma_4 - 3\sigma_2^{-1} \sigma_3^2 - (K_5 + K_6 + K_7) \quad (7.9)$$

where the K_μ are defined by (7.7). Renormalized $|E|^{\text{obs}}$ values $|E'_\mu|$ are defined as

$$|E'_\mu|^2 = |E_\mu|^2 \langle |E_\mu|^2 \rangle^{-1} \\ \approx |E_\mu|^2 [1 - (\langle |E_\mu|^2 \rangle - 1)]. \quad (7.10)$$

Using $(1-x)^{1/2} \approx 1-x/2$, it appears that for equal atoms the K_μ are proportional to the renormalization residues. However, in the distribution they are not linked to the corresponding U values in the main quartet term. It is concluded that non-renormalized $|E|^{\text{obs}}$ values should be used as in the single triplet distribution, with the main quartet terms corrected as in (7.9).

8. Concluding remarks

Applying the central limit theorem for random variables to both the reciprocal vectors and the atomic coordinates, an expression for the joint probability distributions has been derived that can be evaluated by irreducible cluster integrals. The set of multipliers is found by solving maximum-entropy equations, which are equivalent to the multi-channel solution (Bricogne, 1988). The method is valid for unequal scattering factors and for non-uniform dependent *a priori* distributions. The single structure-factor results are consistent with the Debye scattering equation.

The exponential triplet distribution has been derived including interatomic and minimal interatomic distances. The result is in agreement with the Edgeworth exponential triplet distribution for unequal scattering factors up to order $N^{-3/2}$ (Karle & Gilardi, 1973), provided that only $|E|^{\text{obs}}$ values normalized with respect to the total measured reciprocal space are used; renormalization with respect to each reciprocal spherical layer is not correct. The resulting probabilistic triplet inequality may be compared with the Karle-Hauptman triplet inequality. Using a uniform independent *a priori* distribution, the well known exponential quartet distribution (Hauptman, 1975a, b) is found. The cluster integral results indicate that radial interatomic correlations such as interatomic and minimal interatomic distances affect the main quartet exponential term only. Quantitatively these terms can be obtained from the Wilson curve.

APPENDIX A

For the γ_i functions as defined by (3.2) the following identities hold:

$$d\gamma_i(x)/dx = \frac{1}{2} \{ \gamma_{i-1}(x) + \gamma_{i+1}(x) \} - \gamma_i(x) \gamma_i(x) \quad (A.1)$$

$$\gamma_{-i}(x) = \gamma_i(x); \quad \gamma_i(-x) = (-1)^i \gamma_i(x) \quad (A.2)$$

$$\exp [x \cos (\psi)] = I_0(x) \sum_{n=-\infty}^{\infty} \gamma_n(x) \exp (in\psi). \quad (\text{A.3})$$

The following integral is used to evaluate the cluster integrals:

$$\int_{|r|<D} \cos (\mathbf{k} \cdot \mathbf{r}) \, d\mathbf{r} = 4\pi D(nk)^{-2} [\sin (nkD)/nkD - \cos (nkD)]. \quad (\text{A.4})$$

The only terms that have to be taken into account from definitions (4.4) and (4.5) are the terms dependent on both D and k , or possibly invariants. The result for ε_1 follows from definition (4.4) and integral (A.4). The Fourier transform of the first term in ε_1 is equal to the Heaviside function, which is the pair correlation function for a hard-sphere gas $g_0(r)$ in the pair correlation function expansion (Nijboer & Van Hove, 1952):

$$g(r) = g_0(r)\{1 + (N/V)g_1(r) + \dots\} \quad (\text{A.5})$$

$$g_0(r) = H(r-D). \quad (\text{A.6})$$

The result for ε_2 is more difficult to obtain and requires integration over four volumes separated by two intersecting spheres. Integrating over interatomic vectors, the following double integral results in ε_2 (taking $f_\nu = f$ for simplicity):

$$\varepsilon_2 = V\langle\omega\rangle^2 \int_{r_1} \{H(r_1-D) \times \exp [-2\beta f^2 \cos (\mathbf{k} \cdot \mathbf{r}_1)] - 1\} f(\mathbf{r}_1) \, d\mathbf{r}_1 \quad (\text{A.7})$$

where $f(\mathbf{r}_1)$ is again a cluster integral:

$$f(\mathbf{r}_1) = \int_{r_2} \{H(r_2-D) \exp [-2\beta f^2 \cos (\mathbf{k} \cdot \mathbf{r}_2)] - 1\} \times (H(|\mathbf{r}_1 - \mathbf{r}_2| - D) \times \exp \{-2\beta f^2 \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} - 1) \, d\mathbf{r}_2. \quad (\text{A.8})$$

The $f(\mathbf{r}_1)$ is an integral over the four volumes of two intersecting spheres, which can be expressed as

$$f(\mathbf{r}_1) = \int_{r_2} \{\exp [-2\beta f^2 \cos (\mathbf{k} \cdot \mathbf{r}_2)] - 1\} \times (\exp \{-2\beta f^2 \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} - 1) \, d\mathbf{r}_2 + \int_{|r_2|<D, |\mathbf{r}_1 - \mathbf{r}_2|<D} \exp [-2\beta f^2 \{\cos (\mathbf{k} \cdot \mathbf{r}_2) + \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\}] \, d\mathbf{r}_2 - 2 \int_{|r_2|<D} \exp [-2\beta f^2 \cos (\mathbf{k} \cdot \mathbf{r}_2)] \times (\exp \{-2\beta f^2 \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} - 1) \, d\mathbf{r}_2. \quad (\text{A.9})$$

Combining (A.8) with (A.9), taking as contributions only those terms that are dependent on both radii and assuming the intersecting volume to be small relative to each sphere volume, one obtains the

leading term:

$$2V\langle\omega\rangle^2 \int_{|\mathbf{r}_1|<D} \int_{|\mathbf{r}_2|<D} \exp \{-2\beta f^2 \times [\cos (\mathbf{k} \cdot \mathbf{r}_1) + \cos (\mathbf{k} \cdot \mathbf{r}_2)]\} \times (\exp \{-2\beta f^2 \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} - 1) \, d\mathbf{r}_1 \, d\mathbf{r}_2. \quad (\text{A.10})$$

The following approximation in the integrand is made:

$$\exp \{-2\beta f^2 \cos [\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]\} \approx 1 - 2\beta f^2 \cos (\mathbf{k} \cdot \mathbf{r}_1) \cos (\mathbf{k} \cdot \mathbf{r}_2) \quad (\text{A.11})$$

which results in (4.10). Numerical evaluation of the Fourier transform of this result shows that the Fourier transform of the second cluster integral result is almost equal to $g_1(r)$ for a hard-sphere gas (Kirkwood, 1935; Nijboer & Van Hove, 1952):

$$\begin{cases} g_1(r) = \frac{2}{3}\pi(2 - \frac{3}{2}r + \frac{1}{8}r^3) & \text{for } r \leq 2 \\ g_1(r) = 0 & \text{for } r \geq 2. \end{cases} \quad (\text{A.12})$$

The ε_1 and ε_2 results for a single structure factor are in this way found to be consistent with the pair correlation function for a hard-sphere gas by Fourier transform.

APPENDIX B

In order to prove equation (7.8), assume all scattering factors f_ν equal to f , except for a small set of K groups of $2n_\kappa$ atoms ($\kappa \in 1, \dots, K$), in which n_κ atoms have scattering factors equal to $f_\kappa = f + df_\kappa$ and n_κ atoms equal to $f_\kappa = f - df_\kappa$. This ensures that the average of f_ν is equal to f . The df_κ values are all relatively small with respect to f and all n_κ small compared with N .

In that case, evaluation of the σ_n values and using $f_\kappa \approx f$ results in

$$\sigma_2\sigma_4 - \sigma_3^2 \approx 2\sigma_4 \sum_{\kappa} n_\kappa (f_\kappa - f)^2. \quad (\text{B.1})$$

Evaluation of $\langle |F|^2 \rangle = \sum_{\nu} f_\nu^2$ yields in that case

$$\langle |E|^2 \rangle = 1 + 2\sigma_2^{-1} \sum_{\kappa} n_\kappa (f_\kappa - f)^2. \quad (\text{B.2})$$

Combination of (B.1) and (B.2) yields (7.8).

References

- BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 537-548.
 BERTAUT, E. F. (1956). *Acta Cryst.* **9**, 455-460.
 BRICOGNE, G. (1984). *Acta Cryst.* **A40**, 410-445.
 BRICOGNE, G. (1988). *Acta Cryst.* **A44**, 517-545.
 COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
 DEBYE, P. (1915). *Ann. Phys.* **46**, 809-823.
 FELLER, W. (1957). *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd ed. New York: Wiley.
 HAUPTMAN, H. (1975a). *Acta Cryst.* **A31**, 671-679.
 HAUPTMAN, H. (1975b). *Acta Cryst.* **A31**, 680-687.
 HAUPTMAN, H. (1976). *Acta Cryst.* **A32**, 877-882.
 HAUPTMAN, H. & KARLE, J. (1958). *Acta Cryst.* **11**, 149-157.

- JAYNES, E. T. (1957). *Phys. Rev.* **106**, 620-630.
 KARLE, J. (1972). *Acta Cryst.* **B28**, 3362-3369.
 KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* **A29**, 401-407.
 KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181-187.
 KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131-135.
 KIRKWOOD, J. G. (1935). *J. Chem. Phys.* **3**, 300-313.
 KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.
 KRONENBURG, M. J., PESCHAR, R. & SCHENK, H. (1990). *Proc. K. Ned. Akad. Wet.* **93**, 29-32.
 MAYER, J. E. & MAYER, M. G. (1940). *Statistical Mechanics*. New York: Wiley.
- NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* **19**, 734-747.
 NIJBOER, B. R. A. & VAN HOVE, L. (1952). *Phys. Rev.* **85**, 777-783.
 PESCHAR, R. (1987). Thesis, Univ. of Amsterdam, The Netherlands.
 PESCHAR, R. & SCHENK, H. (1987). *Acta Cryst.* **A43**, 513-522.
 SUBRAMANIAN, V. & HALL, S. R. (1982). *Acta Cryst.* **A38**, 577-590.
 VAN KAMPEN, N. G. (1961). *Physica (Utrecht)*, **27**, 783-792.
 WATSON, G. N. (1952). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

Acta Cryst. (1991). **A47**, 480-484

On the Estimation of Non-Measured Diffraction Magnitudes

BY G. CASCARANO, C. GIACOVAZZO AND A. GUAGLIARDI

*Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche, CNR,
c/o Dipartimento Geomineralogico, Campus Universitario, 70124 Bari, Italy*

AND N. STEADMAN

Chemical Crystallography Laboratory, 9 Parks Road, Oxford OX1 3PD, England

(Received 29 June 1990; accepted 26 February 1991)

Abstract

Shortage or ambiguity in diffraction data may limit the efficiency of structure analysis techniques. The joint probability distribution method has been used in order to estimate the values of non-measured diffraction magnitudes. Some experimental tests show the formulae have an efficiency which is promising.

1. Introduction

Sometimes the number of measured diffraction magnitudes is not sufficient for the satisfactory attainment of a crystal structure solution and refinement. Shortage of data occurs regularly in macromolecular crystallography (too many structural parameters to determine compared with the number of available independent observations), but occasionally it also occurs in single-crystal small-molecule crystallography when the diffracting crystal is of poor quality or is unstable under experimental conditions. Shortage of data is very critical in powder crystallography, where occasional or systematic overlapping of diffraction effects does not allow experimental measurement of tens or hundreds of single diffraction amplitudes.

The most effective way for overcoming the problem is to optimize and/or simultaneously to use some complementary experimental techniques (e.g. low-temperature apparatus or synchrotron radiation or a combination of X-ray and/or neutron and/or electron diffraction *etc.*). Often such techniques are not immediately available and the practical problem may

be that of predicting the values of some non-measured diffraction amplitudes by exploiting the measured data as prior information. For example, if data up to $(\sin \theta)/\lambda = \rho_1$ are available, one may try to predict the amplitudes in the range $\rho_2 - \rho_1$ with $\rho_2 > \rho_1$. The problem may be of great importance for crystal-structure-solution methods. In Patterson techniques such supplementary information can make the deconvolution of the Patterson map easier. In direct methods it will make the phasing process and the identification of the correct solution more efficient.

The simplest way to predict the value of a non-measured intensity with vectorial index \mathbf{h} is to use Wilson's statistics. Expectations are

$$\langle |E_{\mathbf{h}}| \rangle = (2/\pi)^{1/2} \quad \text{for centric structures} \quad (1a)$$

$$\langle |E_{\mathbf{h}}| \rangle = (\pi/2)^{1/2} \quad \text{for non-centric structures.} \quad (1b)$$

Relations (1a, b) are too poor to be used for most practical purposes. Probabilistic expressions for estimating $|E_{\mathbf{h}}|$ from all the most reliable quartets in which \mathbf{h} is a cross term were presented by Van der Putten, Schenk & Tsoucaris (1982). More recently, David (1987) suggested a formula which relies on the fact that a Patterson function $P(\mathbf{u})$ is a positive function as well as $P^2(\mathbf{u})$: Sayre's (1952) squaring method was then applied. David's conclusive formula is

$$\langle |F_{\mathbf{h}}|^2 \rangle = \sum_{\mathbf{k}} |F_{\mathbf{k}}|^2 |F_{\mathbf{h}-\mathbf{k}}|^2. \quad (2)$$

Even if of large interest, (2) suffers from two limitations: