gren & Thomas (1982) report on the application of the cumulant expansion (1) '... that such refinements have not converged successfully, rather than that they have not been tried. Our refinement of ten γ values for hydrogen and four for oxygen in the water molecule of $K_2C_2O_4 . H_2O$ resulted in large standard deviations and large correlations. The results of the refinement were not considered meaningful'. Finally, Johnson (1980) himself has noticed deficiencies of the cumulant expansion and locates them primarily in the 'interactions among even-order (real) and odd-order (imaginary) cumulant coefficients'. As a way out, Johnson has established the α formalism with which several structures have been successfully refined in recent years including structures containing mobile atoms in ionic conductors such as AgI (Cava, Reidinger & Wuensch, 1977). Compared to the cumulant expansion, the α formalism differs primarily in that the exponential form of the temperature factor is discarded and the anharmonic terms are expanded into a series of quasi-Hermite polynomials. There is an unequivocally determined inverse Fourier transform to the α formalism that can serve as a p.d.f.; see also Zucker & Schulz (1982). These results from the literature may have their cause in the deficiency of the cumulant expansion that has been pointed out in this paper.

The experimental evidence reported here may not be conclusive for ruling out the cumulant expansion as a useful description of anharmonic motions.

However, we recommend controlling the results obtained with the cumulant expansion by means of corresponding results obtained with other anharmonic expansions.

References

CAVA, R. J., REIDINGER, F. & WUENSCH, B. J. (1977). *Solid State Commun.* **24**, 411–416.
ERIKSSON, A., HERMANSSON, K., LINDGREN, J. & THOMAS, J. O. (1982). *Acta Cryst.* **A38**, 138–142.
International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press (Present distributor: D. Reidel, Dordrecht).
JOHNSON, C. K. (1969). *Acta Cryst.* **A25**, 187–194.
JOHNSON, C. K. (1970). *Thermal Neutron Diffraction*, pp. 132–160. Oxford Univ. Press.
JOHNSON, C. K. (1980). *Computing in Crystallography*, pp. 14.01–14.19. Bangalore: The Indian Academy of Sciences.
KENDALL, M. G. & STUART, A. (1969). *The Advanced Theory of Statistics*. Vol. I. London: Griffin.
KUHS, W. F. (1983). *Acta Cryst.* **A39**, 148–158.
MARCINKIEWICZ, J. (1938). *Math. Z.* **44**, 612–618.
MOSS, B., MCMULLAN, R. K. & KOETZLE, T. F. (1980). *J. Chem. Phys.* **73**, 495–508.
SCHERINGER, C. (1977). *Acta Cryst.* **A33**, 879–884.
ZUCKER, U. H. & SCHULZ, H. (1982). *Acta Cryst.* **A38**, 563–568, 568–576.

Anharmonic Temperature Factors: the Limitations of Perturbation-Theory Expressions

By J. K. MACKENZIE AND S. L. MAIR

CSIRO Division of Chemical Physics, PO Box 160, Clayton 3168, Victoria, Australia

(*Received* 2 *January* 1984; *accepted* 14 *August* 1984)

Abstract

The error in the temperature factor arising from the use of anharmonic perturbation theory is investigated for a one-dimensional one-particle potential by comparison with an exact evaluation of the Fourier transform of the Boltzmann probability density. Results are presented for a range of values of cubic and quartic coefficients for temperature factors derived from moment or cumulant expansions about the harmonic probability density function. It is found that either expansion provides an adequate approximation to the anharmonic temperature factor for moderately small anharmonicity but that both expansions become increasingly inadequate as the anharmonicity gets larger.

1. Introduction

For temperatures where classical approximations are adequate, the starting point for the derivation of the anharmonic temperature factor is the Boltzmann one-particle probability density function (p.d.f.), $\exp[-V(u)/k_BT]$, where $V(u)$ is the one-particle potential (OPP), u is the atomic displacement from a reference position, k_B is the Boltzmann constant and T is the absolute temperature. Expansions about the harmonic (Gaussian) p.d.f. are then made as

moment expansions (*e.g.* Dawson, Hurely and Maslen, 1967; Kurki-Suonio, Merisalo & Peltonen, 1979; Scheringer, 1984*a*). Approximate expressions for the temperature factor are then found by evaluating the ensemble average $\langle \exp i\mathbf{Q} \cdot \mathbf{u} \rangle$ for the corresponding approximate distribution function; here $\mathbf{Q}$ is the scattering vector. The cumulant expansion (Mair, 1980) may be derived from these results.

For either the moment or the cumulant expansions the temperature factors are approximations to the exact temperature factor given by the Fourier transform (FT) of an 'exact' Boltzmann p.d.f. defined hereafter by an OPP expanded to the fourth degree in **u**. In order to examine the nature of this approximation the exact and approximate temperature factors have been evaluated (§§ 2 and 3) for an OPP with cubic and quartic anharmonicity. Since the exact temperature factor must be obtained numerically, attention has, for illustrative purposes, been restricted to a one-dimensional p.d.f. The effect of the perturbation approximation is then discussed in § 4.

2. Exact and approximate OPP temperature factors

The magnitude of the errors in the various approximations is calculated for a simple one-dimensional p.d.f. involving only cubic and quartic anharmonicity.

When the potential

$$V(u) = au^2/2 + bu^3 + cu^4$$

is expressed in dimensionless units by means of the transformation $u = (k_B T/a)^{1/2}x$, the exact Boltzmann p.d.f. becomes

$$p(x) = \exp(-x^2/2 - rx^3 - sx^4)/Z, \qquad (1)$$

where $r = b(k_B T/a^3)^{1/2}$ and $s = ck_B T/a^2$ measure the strength of the anharmonicity and are respectively of order λ and λ^2 in the Van Hove ordering parameter (see Mair, 1980).* The (rescaled) partition function

$$Z = \int_{-\infty}^{\infty} \exp(-x^2/2 - rx^3 - sx^4) \, dx \qquad (2)$$

is a constant that normalizes $p(x)$ so that it is a proper probability density function. The exact temperature factor corresponding to this p.d.f. is

$$T(Q) = T(q) = \int_{-\infty}^{\infty} p(x) \exp(iqx) \, dx \qquad (3)$$

$$= \int_{-\infty}^{\infty} p(x) \cos qx \, dx$$

$$+ i \int_{-\infty}^{\infty} p(x) \sin qx \, dx \qquad (4)$$

---

* We agree with Scheringer [1984*a*, Appendix 4 (supplementary publication No. SUP39686)] that the cumulant expansions in Mair (1980, equations 8, 9, and 14) are not *exact* FT's of the Boltzmann p.d.f. They are, however, approximations to the exact FT correct to order λ^2.

for $q = (k_B T/a)^{1/2}Q$. If the probability density is expressed in terms of the original variable u the value of Z in (1) must be replaced by $Z_{act} = (k_B T/a)^{1/2}Z$. The density function $p(x)$ is well defined if s is greater than zero (or if $r = s = 0$) and provided this is so all the above integrals are convergent and can be evaluated numerically.

The so-called moment expansion for $T(q)$ is derived by expanding $\exp(-rx^3 - sx^4)$ in powers of x, collecting terms for like powers of λ and performing term by term integration in the integrals of (4). To be consistent, all terms up to the same maximum order in λ should be included. Several examples occur in the literature (*e.g.* Willis, 1969; Kurki-Suonio *et al.*, 1979; Tanaka & Marumo, 1983), where some terms of order λ^2 have been included whilst others have been omitted. The temperature factor in (8) of Scheringer (1984*a*) is equivalent to the moment expansion taken to order λ in the cubic coefficients of the OPP and to order λ^2 in the quartic coefficients (the term in r^2 does not occur). It is possible that in certain cases the omitted terms may be numerically insignificant, but it would be unwise to make this assumption *a priori.*

For the case considered here, the result to order λ^2 is

$$T_m(q) = N \exp(-q^2/2)[1 + ir \, \text{He}_3(q)$$
$$- s \, \text{He}_4(q) - r^2 \, \text{He}_6(q)/2]$$
$$= N \exp(-q^2/2)[1 + ir \, (q^3 - 3q)$$
$$- s(q^4 - 6q^2 + 3)$$
$$- r^2(q^6 - 15q^4 + 45q^2 - 15)/2], \qquad (5)$$

where

$$N = 1/(1 - 3s + 15r^2/2) = (2\pi)^{1/2}/Z_m.$$

Z_m is the approximate rescaled partition function and the $\text{He}_n(q)$ are Hermite polynomials as defined by Magnus & Oberhettinger (1949) (see also Cramér, 1946, p. 133). The corresponding cumulant expansion is

$$T_c(q) = \exp[-q^2/2 + ir(q^3 - 3q) - s(q^4 - 6q^2)$$
$$+ r^2(9q^4 - 36q^2)/2], \qquad (6)$$

which is finite as q tends to infinity only if $9r^2 < 2s$.

The inverse FT of (5) is

$$p_m(x) = \exp(-x^2/2)[1 - rx^3 - sx^4 + r^2x^6/2]/Z_m, \qquad (7)$$

which always goes negative for some value of $x(= s^{-1/4}$ for $r = 0)$ and so is not a proper p.d.f. The inverse FT of (6), $p_c(x)$, also behaves in this way because, according to a theorem due to Marcinkiewicz (see Lukacs, 1970, and also Scheringer, 1984*b*) it cannot be the FT of any p.d.f. The numerical calculations below show that it goes negative at about the same value of x as $p_m(x)$ does.

Useful parameters for expressing the degree of anharmonicity are the skewness, γ_1, and the kurtosis, γ_2, defined in terms of the moments of $p(x)$ by the relations

$$\gamma_1 = \langle x^3 \rangle / \langle x^2 \rangle^{3/2}, \tag{8}$$

$$\gamma_2 = (\langle x^4 \rangle - 3\langle x^2 \rangle^2)/\langle x^2 \rangle^2. \tag{9}$$

For a harmonic p.d.f. both γ_1 and γ_2 are zero and γ_1 is also zero when r is zero.

3. Numerical results

The results presented here were obtained by numerical integration and a twelve-point Gaussian quadrature formula was found to be adequate. For the purpose of graphical presentation of the results it is convenient to separate the temperature factor into its real and imaginary parts $T^r(q)$ and $T^i(q)$ and the p.d.f. into even and odd contributions $p^e(x)$ and $p^o(x)$. If r is zero $T^i(q)$ and $p^o(x)$ are zero also. The calculated temperature factors and p.d.f.'s are presented in Figs. 1(a)-(h).

Since temperature factors constitute the more or less directly observed data the main concern here is the agreement between the exact and approximate temperature factors. If either of the approximate temperature factors, $T_m(q)$ or $T_c(q)$, agree sufficiently well with the exact $T(q)$, the p.d.f. given by (1) for the corresponding values of r and s may be taken as a good estimate. If the agreement is poor then the perturbation expansion is inadequate and some other method of analysing the temperature factor is required. The agreement between the p.d.f.'s in the figures is of secondary importance.

(a) Pure quartic anharmonicity ($r = 0$, $s \neq 0$)

For the case $s = 0 \cdot 01$, i.e. for small but significant anharmonicity, with kurtosis $\gamma_2 = -0 \cdot 157$, the approximate and exact $T(q)$ effectively coincide. The p.d.f.'s also coincide, except for a barely discernible deviation of either $p_m(x)$ or $p_c(x)$ near $x = 0$. For $x > 3 \cdot 2$ both $p_m(x)$ and $p_c(x)$ become negative, but negligibly so.

Fig. 1(a) is for $s = 0 \cdot 0276$ ($\gamma_2 = -0 \cdot 284$). This value of s corresponds to the room-temperature values of a and c derived by Kontio & Stevens (1982) for an aluminium atom at a site of $\bar{4}3m$ symmetry in a VAl alloy. The model then provides an approximate one-dimensional analogue of the motion of the Al atom in a $\langle 100 \rangle$ direction. In Fig. 1(a), $T_m(q)$ is indistinguishable from $T_c(q)$ and is starting to show a small deviation from $T(q)$. Deviations of $p_m(x)$ or $p_c(x)$ from $p(x)$ also become just noticeable for both small and large x; both $p_m(x)$ and $p_c(x)$ go negative for $x > 2 \cdot 5$.

In Figs. 1(c), (d) the anharmonicity is very large ($s = 0 \cdot 1$, $\gamma_2 = -0 \cdot 463$). There are now large deviations

of both $T_m(q)$ and $T_c(q)$ from $T(q)$, especially at small values of q where $T_m(q)$ and $T_c(q)$ are both greater than unity. In this region, $T_m(q)$ is a poorer approximation to $T(q)$ than is $T_c(q)$. The corresponding p.d.f.'s are much too large at small x and both $p_m(x)$ and $p_c(x)$ are strongly negative for $x > 1 \cdot 8$.

(b) Cubic and quartic anharmonicity ($r \neq 0$, $s \neq 0$)

Fig. 1(e) is for $r = -0 \cdot 05$, $s = 0 \cdot 0276$ ($\gamma_1 = 1 \cdot 55$, $\gamma_2 = 0 \cdot 427$). These values are again chosen to give a one-dimensional analogue of the room-temperature motion of an aluminium atom in the same VAl alloy but now along a $\langle 111 \rangle$ direction.

The exact and approximate curves for the real part of $T(q)$ in Fig. 1(e) are in good agreement but the small imaginary component is noticeably different for the exact and approximate cases near the turning points. The deviations are, however, still within the error bar based on the experimental uncertainties quoted by Kontio & Stevens (1982) for their α, β and γ. The even and odd components of the p.d.f. are shown in Fig. 1(f) for $x > 0$. The deviations observed in $T^i(q)$ are reflected in $p^o(x)$ for $1 \cdot 5 < x < 3 \cdot 5$. Note that for both the moment and cumulant approximations the odd component $p^o(x)$ is greater than the even component $p^e(x)$ for $2 \cdot 2 < x < 3 \cdot 5$, so that for the corresponding range of negative values $p_m(x)$ and $p_c(x)$ are negative.

Fig. 1(g) shows that when r is doubled the disagreement becomes larger, especially for the imaginary component $T^i(q)$, and perturbation theory is becoming inadequate. For this case $T_c(q)$ is in perceptibly poorer agreement with $T(q)$ than is $T_m(q)$. Moreover, since in this case $9r^2 = 0 \cdot 09 > 2s = 0 \cdot 552$, $T_c(q)$ diverges to infinity as q increases, and so $p_c(x)$ is undefined. This phenomenon is discussed by Scheringer (1984b). Again, $p_m(x)$ goes negative in the range $-3 \cdot 1 < x < -2 \cdot 1$. Clearly, a larger value of r would lead to a more dramatic breakdown of perturbation theory.

4. Discussion

The above results show that when the anharmonicity is moderately small both the cumulant expansion form of the temperature factor and the moment expansion form provide an adequate approximation to the anharmonic temperature factor in practical numerical terms. However, these perturbation-theory expansions become increasingly inadequate as the anharmonicity gets larger. The addition of higher-order terms will not necessarily improve the approximation, as the perturbation expansions ultimately diverge (see Scheringer, 1984a, for a discussion of the divergence of the moment expansion). The moment expansion form has an essential singularity at the origin so that the series does not converge for any finite value of the strength of the anharmonic
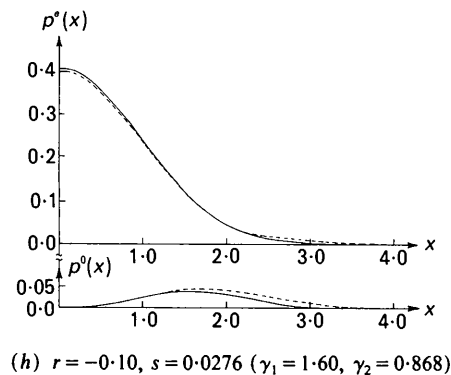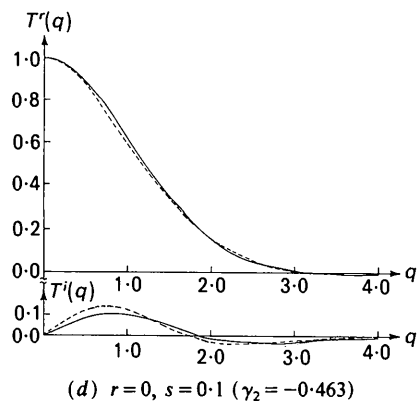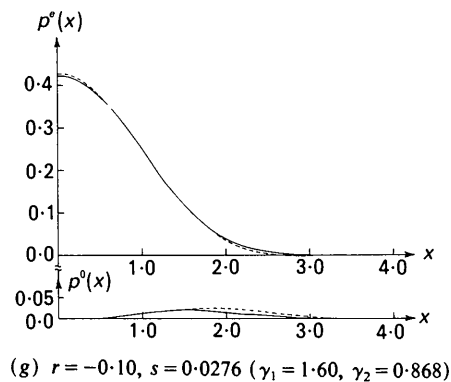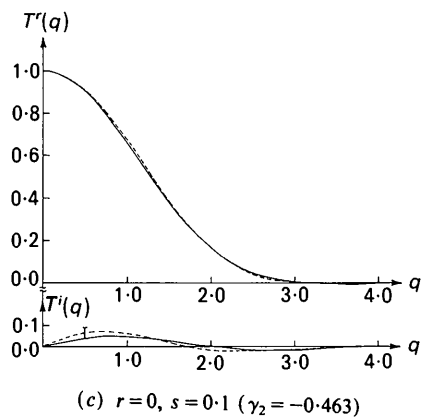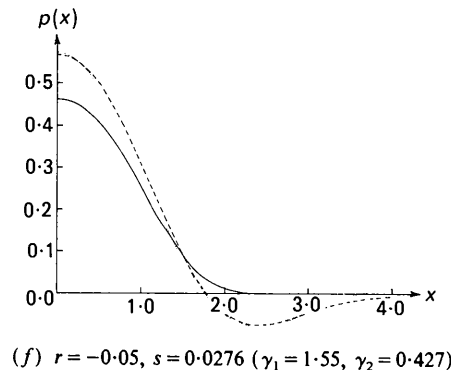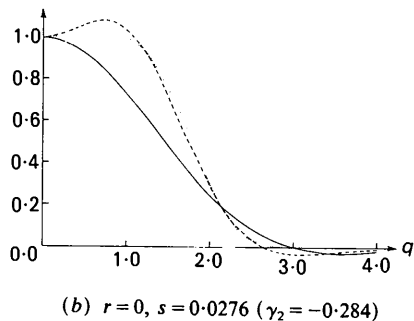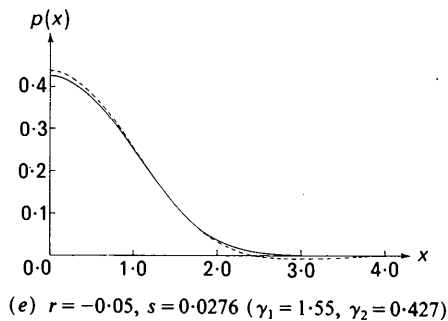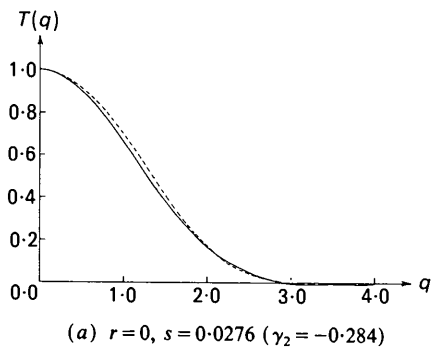
Fig. 1. For the values of the parameters given below each part of the figure, the full line shows either the exact value of $p(x)$ given by equation (1) or its Fourier transform, $T(q)$. The broken line gives the result for the corresponding moment expansion, $T_m(q)$ or $p_m(x)$, while the dotted line gives the result for the cumulant expansion, $T_c(q)$ or $p_c(x)$, if it is distinct from the result for the moment expansion. For (e)-(h) the p.d.f. has even and odd components, $p^e(x)$ and $p^o(x)$, and the corresponding temperature factor has real and imaginary components, $T^r(q)$ and $T^i(q)$.

component of the potential. On the other hand, it is asymptotic in character and when only a finite number of terms are taken the error is less in magnitude than the first term omitted. Thus, although it can be useful numerically for sufficiently weak anharmonicities its accuracy is essentially limited and, for strong anharmonicities, may not be adequate. This is particularly relevant when dealing with data of such high accuracy that deviations of only a few percent in the temperature factor become significant.

Since for the moment expansion, $T_m(q)$, the inverse FT, $p_m(x)$, always exists, one might have hoped that $p_m(x)$ could be used to represent the p.d.f. even if the perturbation approximation were poor, i.e. if $T_m(q)$ differed significantly from $T(q)$. However, the examples considered show that $p_m(x)$ then has regions where it takes significantly negative values despite the fact that it is supposed to be approximating a positive function.

Similar considerations apply to the cumulant expansion. It can be a useful approximation but because of the theorem due to Marcinkiewicz (see Lukacs, 1970) it cannot be the Fourier transform of any p.d.f.; specifically, it must always be positive and therefore cannot follow through a zero of the temperature factor to the negative values that must occur somewhere. Another mathematical difficulty relating to this zero occurs because the cumulant form arises on taking the logarithm of the moment form and so the series for the cumulant form cannot possibly

converge for values of q beyond the zero where the logarithm has its singularity.

The numerical results do not allow a clear choice between the moment and the cumulant expansions. Significant differences between the two approximations appear only when perturbation theory is starting to break down. Under these circumstances a procedure that does not rely on perturbation theory is desirable. If one wishes to retain the OPP approach a suitable procedure would be to express the temperature factor directly as the FT of the exact Boltzmann distribution in accordance with equations (1), (2) and (3). Such computations are well within the capabilities of modern large computers.

References

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.
DAWSON, B., HURLEY, A. C. & MASLEN, V. W. (1967). *Proc. R. Soc. London Ser. A,* **298**, 289-306.
KONTIO, A. & STEVENS, E.D. (1982). *Acta Cryst.* A38, 623-629.
KURKI-SUONIO, K., MERISALO, M. & PELTONEN, H. (1979). *Phys. Scr.* **19**, 57-63.
LUKACS, E. (1970). *Characteristic Functions,* 2nd. ed, p. 213. London: Griffin.
MAGNUS, W. & OBERHETTINGER, F. (1949). *Formulas and Theorems for the Special Functions of Mathematical Physics* (Engl. transl.), p. 80. New York: Chelsea Publishing Co.
MAIR, S. L. (1980). *J. Phys. C,* **13**, 1419-1425.
SCHERINGER, C. (1984a). *Acta Cryst.* A41, 73-79.
SCHERINGER, C. (1984b). *Acta Cryst.* A41, 79-81.
TANAKA, K. & MARUMO, F. (1983). *Acta Cryst.* A39, 631-641.
WILLIS, B. T. M. (1969). *Acta Cryst.* A25, 277-300.

The Estimation of Average Molecular Dimensions.
2.* Hypothesis Testing with Weighted and Unweighted Means

By ROBIN TAYLOR AND OLGA KENNARD

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

Abstract

The average value (μ) of a molecular dimension may be estimated by a weighted ($\bar{x}_w$) or unweighted ($\bar{x}_u$) mean. Computer simulations show that $\bar{x}_u$ can be used in hypothesis tests, since the distribution of $(\bar{x}_u - \mu)/\sigma(\bar{x}_u)$ is closely approximated by Student's t distribution. In contrast, hypothesis tests based on the weighted mean are inexact and potentially misleading.

---

* Part 1: Taylor & Kennard (1983).

I. Introduction

In a previous paper (Taylor & Kennard, 1983) we discussed some of the problems involved in estimating average molecular dimensions from crystallographic data. The average value of k observations of a molecular dimension (x_i, $i = 1, 2, \ldots, k$) is usually estimated in one of two ways. The simplest procedure is to calculate the unweighted mean, $\bar{x}_u$:

$$x_u = \sum_{i=1}^{k} x_i/k, \tag{1}$$

© 1985 International Union of Crystallography