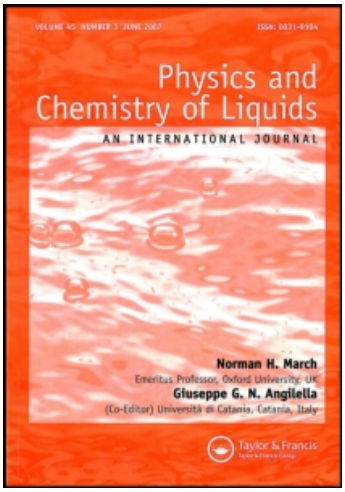

Physics and Chemistry of Liquids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713646857

Optimization of Results Obtained from Integrals over Poisson Distributed Data I. Optimal Correlation Functions from Elastic Neutron Scattering Data in Simple Liquids

Werner A. Schlup[ab]
[a] IBM Watson Laboratory Columbia University New York, New York [b] IBM Zurich Research Laboratory, Rüschlikon, Switzerland

To cite this Article Schlup, Werner A.(1968) 'Optimization of Results Obtained from Integrals over Poisson Distributed Data I. Optimal Correlation Functions from Elastic Neutron Scattering Data in Simple Liquids', Physics and Chemistry of Liquids, 1: 1, 73 — 91
To link to this Article: DOI: 10.1080/00319106808083788
URL: http://dx.doi.org/10.1080/00319106808083788

N Physics and Chemistry of Liquids. 1968. Vol. 1, pp. 73–91

Printed in Germany

Optimization of Results Obtained from Integrals over Poisson Distributed Data

I. Optimal Correlation Functions from Elastic Neutron Scattering Data in Simple Liquids

WERNER A. SCHLUP*

IBM Watson Laboratory
Columbia University
New York, New York

Received July 13, 1967

It is shown how the counting time at different scattering angles in particle scattering experiments has to be chosen so that a physical quantity given by an integral over Poisson distributed data may be optimally determined; i.e., with the smallest possible uncertainties. The method is applied to the elastic scattering of thermal neutrons by liquid Argon. The optimal angle or time distribution of the measurements for obtaining the total and direct particle correlation functions has been calculated from the known data. The result may be used as a starting point for improved measurements.

1. Introduction

Many questions in the theory of particle scattering, where the number of scattered particles at a given angle is measured by counters lead for the evaluation of physical quantities to the following problem: How should the total available counting time be apportioned over the range of angles under investigation or alternatively, how should the density of scattering angles for fixed counting time per angle be chosen so that a physical quantity h expressed by the integral

$$h = \int_a^b K(q)\, I(q)\, \mathrm{d}q \tag{1.1}$$

over a finite interval of integration (ab) of a variable q (e.g., the scattering angle) can be determined as well as possible. The function $K(q)$ of the integrand presumed is known while $I(q)$ is obtained from the data

* Present address: IBM Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland.

(with a statistical uncertainty) only at the discrete points q_n. The error depends on the number of counts and decreases with increasing counting time; i.e., the time during which the counts have been measured. The statistical error of h decreases also; but, since the contribution of the uncertainty in $I(q)$ is weighted by $K(q)$, the points corresponding to large $K(q)$ have to be measured more exactly for an optimal determination of h. Beside the statistical error there is a discontinuity error or sum error arising from the fact, that the integral has to be evaluated as a sum over the points q_n, where measurement has been done. This sum error can be estimated by the Euler-McLaurin sum formula.

In section 2 the statistical assumptions are formulated and the optimization is defined. It is generalized to integral transforms in section 3 and it is applied in section 4 to the correlation functions from the elastic neutron scattering data of liquid Argon. Section 5 contains some final remarks concerning the derived interparticle potential and the normalization of the data.

2. The Principle of Smallest Error

The statistical assumptions will now be specified. The number of counts $N_n = N(q_n)$ for the discrete points q_n with $a \leqq q_1 < q_2 \ldots < q_n < \ldots < q \leqq b$ are assumed to be (1) statistically independent (for pairs, triplets, etc.) and (2) Poisson distributed

$$p(N_n = l) = \frac{\alpha_n^l}{l!} e^{-\alpha_n} \quad l = 0, 1, 2 \ldots \tag{2.1}$$

These assumptions are justified as long as the measurements are done (1) at different times, and (2) the probability of the occurrence of an event is proportional to the time element (Poisson process).

A linear relation exists between $I(q_n)$ and the counting rate:

$$\frac{N(q_n)}{T(q_n)} = A\big(1 + \tau I(q_n)\big) \tag{2.2}$$

where $T(q_n)$ is the counting time at q_n. The coefficients A and $A \cdot \tau$ are known (nonstochastic) parameters in the case of the elastic neutron scattering. A, apart from geometrical factors, is the incoming neutron flux, which because of its high particle density can be measured very exactly and therefore may be assumed to have a sharp distribution. τ is the

coherent scattering ratio, an atomic property which can be measured under a high degree of accuracy in the solid. In the final remarks the conventional normalization (stochastic A) is discussed.

In order to transform the integral (1.1) to a finite sum, we use the Euler-McLaurin sum formula generalized for nonequidistant points q_n. They are assumed to be generated by a normalized point density function

$$\varrho(q) \geqq 0 \quad \text{for} \quad a \leqq q \leqq b \tag{2.3}$$

and its integral

$$P(q) = \int_a^q \varrho(q)\, \mathrm{d}q$$

with the properties

$$P(b) = 1 \tag{2.4}$$

and

$$P(q_n) = \frac{n - \frac{1}{2}}{N} \qquad n = 1, 2, \ldots, N. \tag{2.5}$$

From the results of appendix A the integral can be well approximated by a sum over a large number of terms:

$$\int_a^b F(q)\, \mathrm{d}q = \frac{1}{N} \sum_{n=1}^{N} \frac{F(q)}{\varrho(q)}\bigg|_{q_n} + \frac{1}{24N^2}\left[\frac{1}{\varrho(q)} \cdot \frac{d}{\mathrm{d}q}\left(\frac{F(q)}{\varrho(q)}\right)\right]_{q=a}^{q=b} + O\left(\frac{1}{N^3}\right). \tag{2.6}$$

By means of (2.6) the integral (1.1) can be expressed as

$$h = \sum_{n=1}^{N} \frac{KI}{N\varrho}\bigg|_{q_n} + \frac{1}{24N^2}\left[\frac{1}{\varrho}\left(\frac{KI}{\varrho}\right)'\right]_a^b + O\left(\frac{1}{N^3}\right) \tag{2.7}$$

where the second and third term vanishes for $N \to \infty$, while the first gives the well known Riemann definition of the integral since $1/N\varrho(q_n) \simeq \Delta q_n$.

To apply (2.7) to the given data $I_1 \ldots I_N$, it is necessary to define the boundary values I and I'. We assume according to a lowest order approximation for the second term $I(a) = I_1$, $I'(a) = (I_2 - I_1)\, Np(q_1)$ and a similar expression for $I(b)$, $I'(b)$ in terms of I_{N-1} and I_N. The consequence of the boundary term in (2.7) will be discussed in appendix C.

In the following we restrict ourselves to the main term (sum), a reliable approximation as long as N is large. The average value for h which is a linear combination of statistically independent terms is given by the

sum of the averages of the single terms; in the limit $N \to \infty$ with the I distribution known over a dense set in (ab), it is given by the integral

$$\bar{h} = \int_a^b K\bar{I}\,\mathrm{d}q. \tag{2.8}$$

This means the average is by definition independent of the interval distribution $\varrho(q)$ as long as all intervals vanish in the limit.

The variance, on the contrary, after replacing the sum for $\overline{\Delta h^2}$ by an integral is

$$\overline{\Delta h^2} = \frac{1}{N}\overline{\Delta h_1^2} + O\left(\frac{1}{N^3}\right) \tag{2.9}$$

where

$$\overline{\Delta h_1^2} = \int_a^b K^2\overline{\Delta I^2}\,\frac{\mathrm{d}q}{\varrho}. \tag{2.10}$$

It depends inversely on the point distribution and on the total number of points. The main term will be the smaller the higher the density of measured points.

In the limit $N \to \infty$ (see appendix B) an asymptotic distribution for h will exist, provided that K, $\bar{I}$, $\overline{\Delta I^2}$ and ϱ are given functions and the semi-invariants of h exist. Under these assumptions the central limit theorem holds, i.e., the standardized variable h has a Gaussian distribution.

We require that the point distribution $\varrho(q)$ be such that the error δh in the quantity h is a minimum and we identify the error with the square root of the variance

$$\delta h = (\overline{\Delta h^2})^{1/2}. \tag{2.11}$$

In neglecting terms of the order $O\left(\frac{1}{N^3}\right)$ we have the following variational problem:

$$\frac{\delta\overline{\Delta h_1^2}}{\delta\varrho} = 0 \tag{2.12}$$

with the normalization condition (2.4). The extremal solution

$$\varrho(q) = \frac{(K\overline{\Delta I^2})^{1/2}}{\int_a^b (K^2\overline{\Delta I^2})^{1/2}\,\mathrm{d}q} \tag{2.13}$$

is a positive function and represents a minimum (see appendix D). The optimal variance is then

$$(\overline{\Delta h^2})_{\text{opt}} = \frac{1}{N}\left(\int_a^b (K^2\overline{\Delta I^2})^{1/2}\,\mathrm{d}q\right)^2. \tag{2.14}$$

It decreases as the number of points N is increased.

If the point distribution is given a priori by the diffraction pattern $\varrho(q) = \varrho_0(q)$, as for the case of equally spaced scattering angles, we may choose the counting time $T(q)$ to optimize h, since, from Eqs. (2.1) and (2.2)

$$\overline{\Delta I^2} = \frac{1 + \overline{\tau I}}{A\tau^2\, T(q)}, \tag{2.15}$$

only the product $T(q) \cdot \varrho(q)$ will be fixed by the optimization under the condition of a given total time T_t

$$T_t = \sum_{n=1}^{N} T(q_n) = N \int_a^b T\varrho\,\mathrm{d}q. \tag{2.16}$$

The time distribution of the measurements is then

$$T \cdot \varrho_0 = \text{Const}\,(K^2(1 + \tau\overline{I})^{1/2}). \tag{2.17}$$

Physically this means: the precision of the result depends only on the product of counting time and point distribution apart from higher terms in the Euler-McLaurin formula. For example, measuring at two neighboring points q_1, q_2 for times T_1, T_2 or measuring only in one intermediate point (e.g., $(q_1 + q_2)/2$), for the total time $T_1 + T_2$ are equivalent procedures. In this case $T(q)$ has been doubled and $\varrho(q)$ halved. The density of points necessary to consider only the main term can be estimated from the Euler-McLaurin formula.

3. Optimization of the Correlation Functions

In the theory of elastic, nonmagnetic scattering of neutrons in liquids the differential cross section is given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = N_0\overline{a^2}(I + \tau\, I(q)) \tag{3.1}$$

where N_0 = number of atoms of the liquid in scatterer
n_0 = particle density in scatterer
a_i = scattering length of i-isotope
$\tau = \overline{a^2}/\overline{a}^2$,

and—means the average over the isotopes present. The particle pair density function

$$n_2(r_1, r_2) = n_0^2(1 + h(r_1 - r_2)) \tag{3.2}$$

defines the total correlation function $h(r)$ which is related to $I(q)$ through a Fourier transform

$$h(r) = \frac{1}{(2\pi)^3 \cdot n_0} \int \mathrm{d}q^3 \, e^{iqr} \, I(q). \tag{3.3}$$

A more direct connection with the interparticle potential has the direct correlation function $c(r)$ introduced by Ornstein, Zernike

$$c(r) = \frac{1}{(2\pi)^3 n_0} \int \mathrm{d}q^3 \, e^{iqr} \frac{I(q)}{1 + I(q)}. \tag{3.4}$$

In approximate theories the potential can be locally expressed in terms of h and c; for large r the simple asymptotic relation

$$\varphi = -kT \cdot c \tag{3.5}$$

holds.

$h(r)$ is now (cf. section 2) a random variable depending on one parameter; (1.1) becomes the integral transform

$$h(r) = \int_a^b K(r, q) \, I(q) \, \mathrm{d}q \tag{3.6}$$

with kernel

$$K(r, q) = \frac{q \cdot \sin qr}{2\pi^2 n_0 r}. \tag{3.7}$$

In order to generalize the concept of optimization we have to find what space average we want to be small. It seems reasonable to take the unweighted space average

$$\widetilde{\overline{\Delta h^2}_r} = \frac{\int d^3r' \cdot \overline{\Delta h^2}(r')}{\int d^3r' \cdot 1} \tag{3.8}$$

where the integration goes over a spherical shell with the thickness R around r:

$$r - R/2 < |r'| < r + R/2 \tag{3.9}$$

If the condition

$$r \gg R \gg \frac{\pi}{q} \tag{3.10}$$

is fulfilled the integration goes over many oscillations with nearly constant amplitude. The average thus becomes independent of R. The averaged variance has the kernel

$$K_0^2 = K^2(r) = \frac{q^2}{8\pi^4 n_0^2 r^2}. \tag{3.11}$$

This gives the optimal point distribution for the determination of h

$$\varrho_h(q) = \frac{q(\overline{\Delta I^2})^{1/2}}{\int_a^b \mathrm{d}q\, q(\overline{\Delta I^2})^{1/2}}. \tag{3.12}$$

The meanvalue and variance of c can be found approximately by assuming that the distribution of I is rather sharp and

$$\overline{\Delta I^2} \gg (1 + \bar{I})^2. \tag{3.13}$$

We can therefore use a Taylor expansion for $\dfrac{\mathrm{I}}{1+\mathrm{I}} - \overline{\left(\dfrac{\mathrm{I}}{1+\mathrm{I}}\right)}$:

$$\overline{\left(\Delta \frac{I}{1+I}\right)^2} = \frac{\overline{\Delta I^2}}{(1+\bar{I})^4}. \tag{3.14}$$

The space averaged variance of c will be

$$\widetilde{\overline{\Delta c^2}} = \frac{1}{N} \int_a^b K_0^2 \frac{\overline{\Delta I^2}}{(1+\bar{I})^4} \cdot \frac{\mathrm{d}q}{\varrho} + O\left(\frac{1}{N^3}\right) \tag{3.15}$$

and the optimal point density for c correspondingly

$$\varrho_c(q) = \frac{q \dfrac{(\overline{\Delta I^2})^{1/2}}{(1+\bar{I})^2}}{\int_a^b q \dfrac{(\overline{\Delta I^2})^{1/2}}{(1+\bar{I})^2} \mathrm{d}q}. \tag{3.16}$$

Introducing the obtained densities gives for the optimal variance

$$\widetilde{\overline{\Delta h^2}}, \widetilde{\overline{\Delta c^2}} \propto \frac{1}{N \cdot r^2}. \tag{3.17}$$

The errors will be the smaller the larger r. Since generally h and c decrease faster than $1/r$ and small r values do not satisfy (3.10), the optimization will be useful in practice only in a range of medium separation.

4. Application of the Optimization to the Correlation Functions of Liquid Argon

For the evaluation of the optimal point density distribution it is necessary to know the exact meanvalue $\bar{I}$ which should have the data normalization property

$$h(o) = -1 = \frac{1}{2\pi^2 n_0} \int_0^\infty \mathrm{d}q\, q^2 \bar{I}(q). \tag{4.1}$$

A series of measurements will give a good meanvalue which, however, does not fulfill the condition (4.1) exactly. But starting from it, we could find a good point distribution which would make a better determination possible. Again, this would improve the optimal point distribution and so on. After a few steps practically the best distribution for a given total time could be obtained.

As a good meanvalue we use the neutron scattering data of Henshaw[1] for liquid Argon at 84 °K and a neutron wavelength $\lambda = 1.04$ Å. The optimization is done with respect to both correlation functions $h(r)$ and $c(r)$, but with the integration over q going from $a = q = 0$ to $b = Q = 7.15$ Å. This assumes an extrapolation of the data to zero wavevector which can be accomplished by using the relation

$$I(o) = n_0 \int \mathrm{d}r^3 h(r) = -1 + kTn_0 \cdot \chi_T \tag{4.2}$$

where χ_T is the isothermal compressibility.

The optimization introduced in the preceding sections can now be calculated explicitly. When $h(r)$ is optimal then the normalized point distribution is given in Figure 1. The density of the points where the measurements should be made is linear with q according to (3.12) and, therefore, also essentially linear in the angle $\theta(\theta/2 \leqq 35°)$, but modulated with $(\overline{\Delta I^2})^{1/2}$ or $(1 + \tau \bar{I}^{1/2})$. ϱ_h will exhibit a maximum near the point where $\bar{I}$ has a sharp maximum. The same is true for the minima. This means in points of high intensity one should measure with higher accuracy (denser points or a longer time) and additionally larger angles have to be determined more exactly in order to be able to derive good $\overline{h(r)}$ from

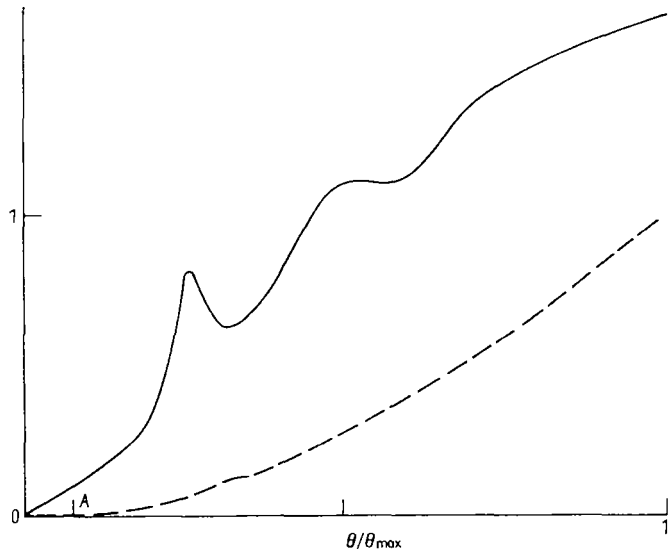

Figure 1. The full curve is the point distribution ϱ and the dashed curve its integral P versus scattering angle θ/θ_{max} in the case of optimal total correlation function $h(r)$. A gives the smallest angle measured.

the data. The points for small q are only of minor importance for the variance and for $q = 0$ one has even a vanishing density. This means the uncertainties arising from small q values are negligible. But, this does not mean that no measurements at all are necessary in a range of small q; a density $q = 0$ in (ab) gives an infinite contribution to the variance.

Figure 2 shows the point density for an optimal $c(r)$. An extremely high maximum is obtained for small q values, since the denominator of (3.14) vanishes almost because of $I(o) \gtrsim -1$. This maximum lies somewhat below the lower cutoff on the measurements.

At the point where $I(q)$ has a maximum ϱ_c exhibits a minimum; for larger q the density ϱ_c is rather constant. If one looks at the function $P(q)$ defined by (2.3), approximately 4/5 of all points should be chosen to measure the range below the first maximum of $I(q)$. Only a few points are then left to measure in the remaining interval. It is therefore necessary to convince oneself that the sum error in (2.6) is still negligible, otherwise higher order terms have to be considered in the Euler-McLaurin formula.

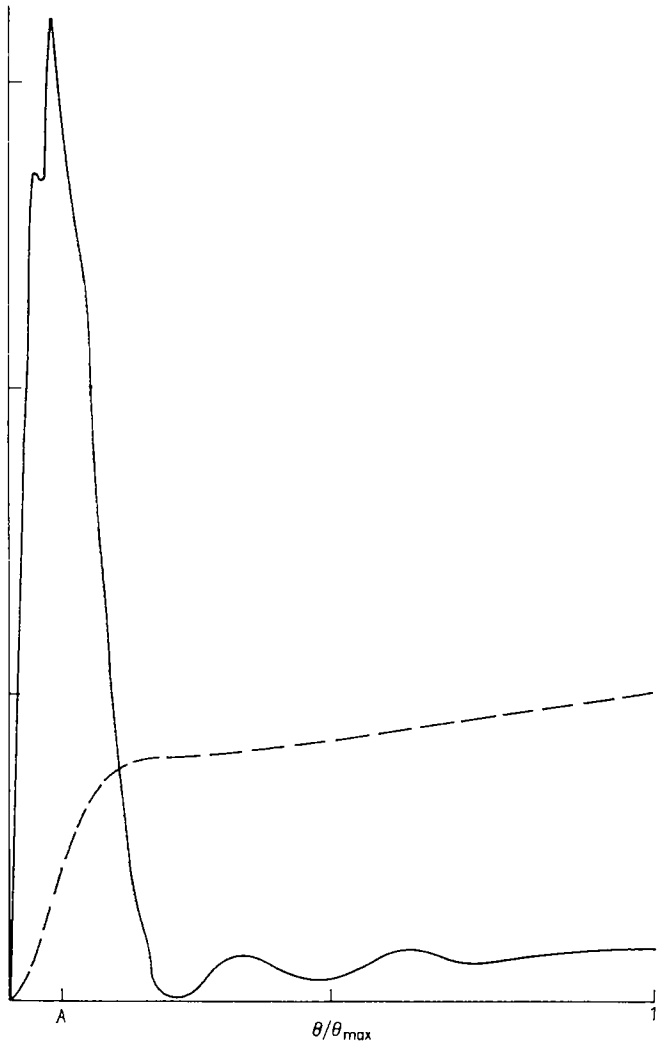

Figure 2. The full curve is the point distribution ϱ and the dashed curve its integral P versus scattering angle θ/θ_{max} in the case of optimal direct correlation function $c(r)$. A gives the smallest angle measured.

The optimal point distribution for h and c are therefore very different and if the incoherent part of the scattering is negligible ($\tau \simeq 1$) the following relation holds

$$\varrho_h^3 \cdot \varrho_c = \text{Const. } q^4. \tag{4.3}$$

This proves that h and c can never simultaneously be optimal. The better c the worse is h and vice versa. In general c is characteristic for the pair potential and the distribution of figure 2 has to be adopted in the experiments. This fact emphasizes the extraordinary importance of $I(q)$ in the range before the first maximum of $I(q)$. Since this region has been covered only partly (~40%) until now by the experiment, we cannot expect great accuracy of the results available at present.

5. Final Remarks

It has been shown that the space average of the correlation functions are optimal for a definite point or counting time distribution. The question is how much better the pair potential can be derived. This will be discussed within the approximate theories of Percus, Yevick[2] (PY) and the hypernetted chains[3] (HNC). In both cases the pair potential can be locally represented by h and c; the meanvalue $\bar{\varphi}$ and the variance $\overline{\Delta\varphi^2}$ can be expressed by the (mixed) variances of h and c. This assumes the validity of a variance expansion like in (3.14) (see reference 4). In both cases the potential is given asymptotically by Eq. (3.5). Therefore it should be expected that an optimal c is of higher importance for an optimal pair potential. The numerical discussion of the Argon data confirms this conjecture. The uncertainties δh and δc show essentially a hyperbolic behavior according to Eq. (3.17). The individual structure of $I(q)$ enters only through small deviations from this behavior which are less than 10% in the range of medium separation (3–10 Å). Therefore the following relations hold approximately

$$\delta h_h \sim 0.9\,\delta h, \quad \delta h_c \sim 2.0\,\delta h$$

$$\delta c_h \sim 3.5\,\delta c, \quad \delta c_c \sim 0.5\,\delta c,$$

where δh, δc are the uncertainties for the equidistant angle measurements and subscripts h and c refer to optimization with respect to h and c. The product of errors $\delta h \cdot \delta c$ is much larger in the case of the h optimization than for the c optimization, where it is about the same as for equal angle spacing. Figures 3 and 4 give the uncertainties for the interparticle potential derived from the HNC and the PY equations, respectively. It can be seen that the uncertainty in the potential reduces by a factor 2 if the optimization is done with respect to the direct correlation function. This increase in accuracy is decisive for a critical discussion of the po-

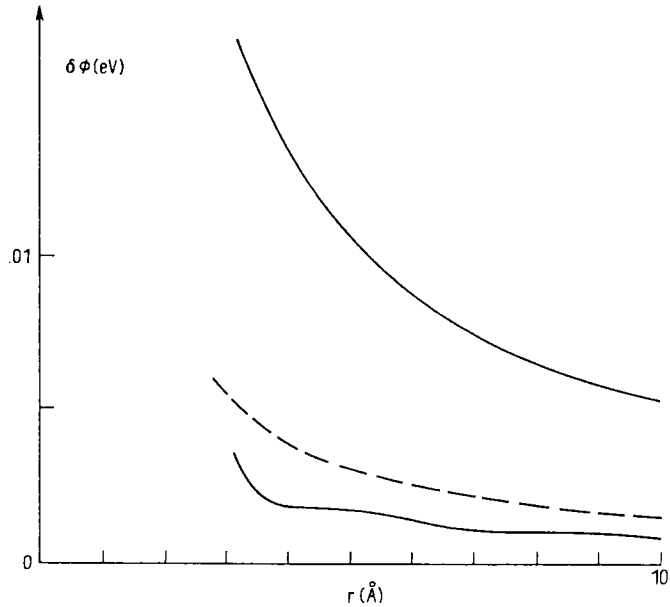

Figure 3. The uncertainties in the potential derived from the *HNC* equation are given for the h optimization (upper curve) and the c optimization (lower curve). The dashed curve belongs to equally spaced angles. The value of the potential depth for Argon is about 0.01 eV.

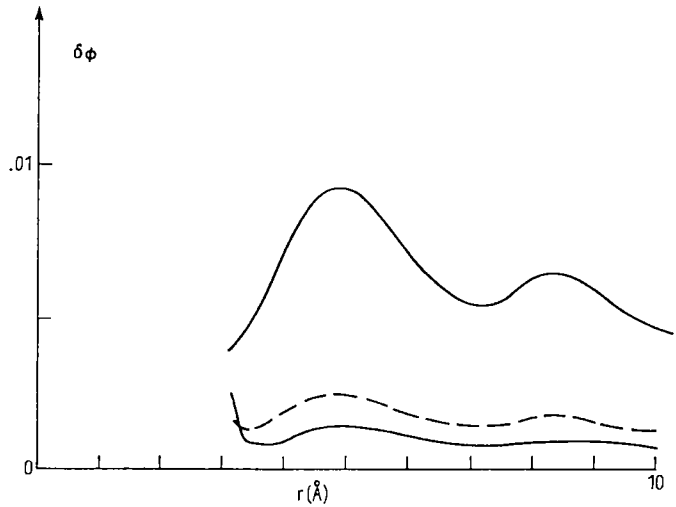

Figure 4. The uncertainties derived from the *PY* equation in the same order as in Figure 3.

tential itself, especially with respect to the Kohn effect[5] for the effective interaction in liquid metals.

An additional statistical error in the evaluation of the data arises from the conventional normalizations by means of $h(o) = -1$ for every set of measurements. As stated earlier, it can be avoided by a highly accurate measurement of the incoming neutrons. In the former case the normalization constant A will be a stochastic variable which depends linearly on the N_n. The intensity I_n is then a nonlinear function of the N_n. The mean value of I_n can be expanded in a variance expansion which contains additional terms of the order $O(1/N)$ arising from the correlation of the normalization constant with N_n (see appendix E). As long as the error in the normalization $\overline{\Delta h^2(o)}$ is small, we can use the variance expansion for $h(r)$ with stochastic normalization, If, however, $\overline{\Delta h^2(o)}$ is large the error in the normalization influences strongly the conventional evaluation of the data compared to our method.

Acknowledgement

My thanks are due to Drs. M. C. Gutzwiller and S. H. Koenig for helpful discussions and critical reading of the manuscript.

Appendix A

The Euler-McLaurin sum formula[6] for the same distance of points and $0 \leqq \alpha \leqq 1$, $M < N$ (M, N integers) and L-times continuously differentiable functions $f(x)$ is

$$\int_M^N f(x)\,dx = \sum_{n=M}^{N} f(n+\alpha-1) - \sum_{l=1}^{L} \frac{B_l(\alpha)}{l!}\left(f^{(l-1)}(N) - f^{(l-1)}(M)\right) + R_L \tag{A.1}$$

with the residual term

$$R_L = \frac{(-)^{L-1}}{L!}\int_M^N B_L^*(x-\alpha)\, f^{(L)}(x)\,dx. \tag{A.2}$$

$B_l(x)$ are the Bernoulli polynomials defined by application of (A.1) to $e^{\alpha x}$ for $L \to \infty$

$$\frac{te^{xt}}{e^t - 1} = \sum_{l=0}^{\infty} \frac{B_l(x)}{l!}\, t^l \tag{A.3}$$

and

$$B_l^*(x) = B_l(x - [x]) \tag{A.4}$$

is the $B_l(x)$ in $0 \leqq x \leqq 1$ and its periodic continuation with period $= 1$.

(A.1) allows us to express an integral as (I) a sum over contributions of equidistant points $x_n = n + \alpha - 1$, (II) the boundary values of the function and its derivatives, and (III) the residual term which is an integral over the whole range (ab). If one applies (A. 1) to smooth, slowly varying functions, one can approximate the integral by the sum and vice versa. The sum error which arises consists of the contributions (II) and (III). In general (III) does not vanish for $L \to \infty$ and therefore the sum of (II) is only semiconvergent; this is obvious when $N \to +\infty$, $M \to -\infty$ and the function and all the derivatives vanish at the boundaries (e.g., $\bar{e}^{x^2}$).

In order to generalize the sum formula (A.1), we introduce a continuous density function ϱ whose L derivatives are all continuous with the following properties:

$$\varrho(q) > 0 \text{ in } (ab), \int_a^b \varrho \, dq = 1. \tag{A.5}$$

$\varrho(q)$ generates a nonequidistant series of points q_n through

$$x(q) = N \int_a^q \varrho(q') \, dq' \tag{A.6}$$

and

$$x(q_n) = x_n = n + \alpha - 1 \quad n = 1, 2, \ldots, N.$$

(A.1) is then

$$\int_a^b F(q) \, \varrho(q) \, dq = \frac{1}{N} \sum_{n=1}^{N} F(q_n) - \sum_{l=1}^{L} \frac{B_l(\alpha)}{N^l \cdot l!} \left[\left(\frac{d}{\varrho(q) \, dq} \right)^{(l-1)} F(q) \right]_{q=b}^{q=a} + R_L \tag{A.7}$$

or for a density independent integrand $g(q)$

$$\int_a^b g(q) = \sum_{n=1}^{N} \frac{g(q_n)}{N \varrho(q_n)} - \sum \frac{B_l(\alpha)}{N^l l!} \left[\left(\frac{d}{\varrho(q) \, dq} \right)^{(l-1)} \frac{g(q)}{\varrho(q)} \right]_{q=a}^{q=b} + R_L. \tag{A.8}$$

The residual term will be

$$R_L = \frac{(-)^L}{L! \, N^L} \int_a^b B_L^*(x(q) - \alpha) \left(\left(\frac{d}{\varrho(q) \, dq} \right)^{(L)} \frac{g(q)}{\varrho(q)} \right) \varrho(q) \, dq. \tag{A.9}$$

(A.8) and (A.9) represent the general one dimensional Euler-McLaurin sum formula. It consists again of the three parts $\int = (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III})$; (I) is the sum over the contributions in nonequidistant points q_n, (II) gives an analytic expression for the sum error whose last term is of the same order (in N) as the residual term (III), which can be estimated.

The most convenient choice for α is $B_1(\alpha) = 0$ where the function itself does not appear in (II) and further all even terms in (II) vanish. The lowest nontrivial form of (A.1) is given by $L = 3$

$$\int_a^b g(q)\,\mathrm{d}q = \sum_{n=1}^{N} \frac{g(q)}{N\varrho(q)}\bigg|_{q_n} + \frac{1}{24N^2}\left[\frac{1}{\varrho}\left(\frac{g}{\varrho}\right)'\right]_{q=a}^{q=b} + R_3 \qquad (A.10)$$

$$R_3 = \frac{1}{6N^3}\int_a^b B_3^*\left(x(q) - \frac{1}{2}\right)\left(\left(\frac{d}{\varrho\,\mathrm{d}q}\right)^{(3)}\frac{g}{\varrho}\right)\varrho\,\mathrm{d}q. \qquad (A.11)$$

Because of

$$|B_3^*(\alpha)| \leqq \frac{\sqrt{3}}{36}$$

the residual term can be estimated to be

$$|R_3| \leqq \frac{\sqrt{3}}{216}\cdot\frac{1}{N^3}\int \mathrm{d}q\left|\frac{d}{\mathrm{d}q}\cdot\frac{d}{\varrho\,\mathrm{d}q}\cdot\frac{d}{\varrho\,\mathrm{d}q}\cdot\frac{g}{\varrho}\right|. \qquad (A.12)$$

It is of lower order $O\left(\frac{1}{N^3}\right)$ and can be neglected for large N compared to the second order derivative term. It should be mentioned that the largeness of N is mainly determined by the smoothness of the functions $g(q)$ and $\varrho(q)$.

Appendix B

We want the limit distribution for h, when the number of points $N \to \infty$, but where the point distribution shall be assumed to be fixed. Further it is assumed that the central moments exist for every N that can be proved, as long as $\varrho \neq 0$ in (ab). A first order zero of $\varrho(q)$ at the boundary $q = 0$ arising for example from the optimization is harmless since $K(q)$ has then a second order zero.

The logarithm for the characteristic function $\varphi_h(t)$ of h is

$$\ln\left(\overline{e^{ith}}\right) = \ln\varphi_h(t) = \psi_h(t) = \sum_{l=1}^{\infty}\frac{(\mathrm{i}t)^l}{l!}\overline{Sh^l} \qquad (B.1)$$

where $\overline{Sh^l}$ are the semi-invariants for the random variable h. The average $\bar{h} = \overline{Sh^1}$, and the higher $\overline{Sh^l}$ can be expressed by the central moments only. Therefore

$$\psi_{h-\bar{h}}(t) = \sum_{l=2}^{\infty} \frac{(\mathrm{i}t)^l}{l!} \overline{Sh^l}. \tag{B.2}$$

h is after Eq. (2.7) a linear combination of statistically independent terms

$$h = \sum_{n=1}^{N} a_n I_n \tag{B.3}$$

whose coefficients in lowest order in N are given by

$$a_n = \frac{K(q_n)}{N\varrho(q_n)} - \frac{K}{24N\varrho}\bigg|_b \delta_{nN-1} + \frac{1}{24}\left(\frac{K}{N\varrho} + \frac{1}{N^2\varrho}\left(\frac{K}{\varrho}\right)'\right)\bigg|_b \delta_{nN} \tag{B.4}$$

plus a corresponding contribution for the lower boundary. Since the characteristic function of independent variables have to be multiplied, the semi-invariants for h and I_n are related through

$$\overline{Sh^l} = \sum_{n=1}^{N} a_n^l \, \overline{SI_n^l}. \tag{B.5}$$

In transforming the sum to an integral retaining only terms of lowest order (besides the integral) we find

$$\overline{Sh^l} = \int_a^b \frac{K^l \overline{SI^l}}{N^{l-1}\varrho^{l-1}} \,\mathrm{d}q + \frac{l-1}{24N^{l+1}\varrho}\left(\frac{K^l\overline{SI^l}}{\varrho^l}\right)'\bigg|_a^b + \ldots, \tag{B.6}$$

the integral term is $O\left(\frac{1}{N^{l-1}}\right)$ and the next of two orders higher. For $l = 1$ all correction terms vanish and we get Eq. (2.8).

The Poisson distribution (2.1) has the semi-invariants

$$\overline{SI_n^l} = \frac{\alpha_n}{(A\tau)^l} = \frac{A(1+\tau\bar{I})}{(A\tau)^l}. \tag{B.7}$$

From (B.6) and (B.7) follows that the semi-invariants Sh^l even exist for $a = 0$ and $\varrho(q) \approx q$, $K(q) \approx q^2$.

We now consider the limit $N \to \infty$ for the variable $\sqrt{N(h-\bar{h})}$; because of

$$\psi_{\sqrt{N(h-\bar{h})}}(t) = \psi_{h-\bar{h}}\left(\sqrt{N}\cdot t\right)$$

only the second order term in (B.2) remains:

$$\lim_{N\to\infty} \psi_{\sqrt{N(h-\bar{h})}}(t) = -\frac{t^2}{2} \cdot \int_a^b \frac{K^2 \overline{\Delta I^2}}{\varrho} \, dq. \tag{B.8}$$

The standardized variable

$$z = \sqrt{N} \frac{h - \bar{h}}{(\overline{\Delta h_1^2})^{1/2}} \tag{B.9}$$

has a Gaussian distribution for $N \to \infty$:

$$f(z) = \frac{1}{\sqrt{2\pi}} \bar{e} \frac{z^2}{2}. \tag{B.10}$$

There are similar limit theorems for pairs of h values at different points in the space. Because of the correlation of $h(r)$ and $h(r')$ one finds for the standardized variables z and z' a correlated Gaussian in the limit $N \to \infty$.

Appendix C

Apart from terms of $O(1/N^4)$ the variance $\overline{\Delta h^2} = \overline{Sh^2}$ follows from (B.6)

$$\overline{\Delta h^2} = \int_a^b \frac{K^2 \overline{\Delta I^2}}{N\varrho} dq + \frac{1}{24N^3\varrho} \left(\frac{K^2 \overline{\Delta I^2}}{\varrho^2} \right)' \Bigg|_a^b. \tag{C.1}$$

Thereby the boundary values have been constructed from the secant of the points $n = 1,2$ and $n = N - 1, N$.

The variation of the space average of $\overline{\Delta h^2}$ under the condition

$$\int_a^b \varrho \, dq = 1 \tag{C.2}$$

gives the well known solution with the additional results

$$\varrho(a) = \varrho(b) = \infty \tag{C.3}$$

in order to make the variation vanish at the boundaries $q = a, b$. Since the normalization should be fulfilled, the singularity in ϱ should have a vanishing measure and would not contribute to the transformation $x(q)$ at all.

The nature of the singularity will be obvious if one writes the boundary terms in (C.1) under the integral with a delta function exhibiting a finite width of the order of the distance $q_2 - q_1$ or $q_N - q_{N-1}$. This is consistent with the derivation which defines the boundary from the properties in the interior points $n = 1,2$ or $n = N - 1, N$. If the width would vanish independently of N one could verify the result (C.3) stated above.

Appendix D

The variational problem in simplified notation is

$$S = \int_a^b \frac{A(q)}{\varrho(q)}\,\mathrm{d}q, \quad A(q) \geqq 0 \tag{D.1}$$

$$\frac{\delta S}{\delta \varrho} = 0 \quad \text{and} \quad \int_a^b \varrho(q)\,\mathrm{d}q = 1. \tag{D.2}$$

The solution is

$$\varrho_0(q) = \frac{\sqrt{A(q)}}{\int_a^b \sqrt{A(q)}\,\mathrm{d}q} \geqq 0 \tag{D.3}$$

$$S_0 = \Big(\int_a^b \sqrt{A(q)\,\mathrm{d}q}\Big)^2. \tag{D.4}$$

The second variation in the solution is

$$\delta^2 S = \int_a^b \frac{2A(q)}{\varrho^3(q)}\,(\delta\varrho(q))^2\,\mathrm{d}q \tag{D.5}$$

which is always positive for small departures from ϱ_0.

Appendix E

With the conventional normalization of the data ($h(o) = -1$) the correlation function can be written

$$h(r) = -\alpha(r) + \beta \frac{\sum_{n=1}^{N} A_n(r)\,N_n}{\sum_{n-1}^{N} A_n(o)\,N_n} \frac{A(r)}{B} \beta - \alpha(r) \tag{E.1}$$

where

$$\alpha(r) = \frac{1}{\tau} \sum_{n=1}^{N} A_n(r),$$

$$\beta = \alpha(o) - 1, \tag{E.2}$$

and $A_n(r)$ is essentially the kernel of the Fourier transform. We assume

$$O\left(\frac{1}{N}\right) = \overline{\Delta B^2} \ll \overline{B}^2 = O(1) \tag{E.3}$$

which can be accomplished when N is large enough; (E.3) is well satisfied for Henshaw's data. Then again we can expand with respect to the variances of A and B

$$\overline{h}(r) = -\alpha(r) + \beta\left(\frac{\overline{A}}{\overline{B}} - \frac{1}{\overline{B}^2}\overline{\Delta A \cdot \Delta B} + \frac{\overline{A}}{\overline{B}^3}\overline{\Delta B^2}\right)$$

$$\overline{\Delta h^2}(r) = \beta^2\left(\frac{\overline{\Delta A^2}}{\overline{B}^2} - \frac{2\overline{A}}{\overline{B}^3}\overline{\Delta A \cdot \Delta B} + \frac{\overline{A}^2}{\overline{B}^4}\overline{\Delta B^2}\right). \tag{E.5}$$

The first terms give the contribution if the normalization is assumed to be sharp. The higher terms give corrections which generally for small r decrease the uncertainty. For $r = 0$, $A = B$ and per definition $\overline{h(0)} = 1$ and $\overline{\Delta h^2(0)} = 0$.

References

1. D. G. Henshaw, *Phys. Rev.* **105**, 976 (1957).
2. I. K. Percus, G. J. Yevick, *Phys. Rev.* **110**, 1 (1958).
3. G. S. Rushbrooke, *Physica* **26**, 259 (1960).
4. W. A. Schlup, Sumposium on Inelastic Scattering of Neutrons by Condensed Systems p. 183 (Brookhaven National Laboratory, 1965, BNL 940 (C-45)).
5. W. Kohn, S. H. Vosko, *Phys. Rev.* **119**, 912 (1960).
6. M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions, p. 806 (National Bureau of Standards, 1964).