A comparison of different Fourier transform procedures for analysis of diffraction data from noble gas fluids ⊘

J. E. Proctor ✉ 🆔 ; C. G. Pruteanu 🆔 ; B. Moss 🆔 ; M. A. Kuzovnikov 🆔 ; G. J. Ackland 🆔 ; C. W. Monk 🆔 ;
S. Anzellini 🆔

Check for updates

View Online   Export Citation   CrossMark

Articles You May Be Interested In

Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide *Q*-range

J. Chem. Phys. (February 2013)

A comparison of different Fourier transform procedures for analysis of diffraction data from noble gas fluids

View Online   Export Citation   CrossMark

J. E. Proctor,[1,a] C. G. Pruteanu,[2] B. Moss,[1] M. A. Kuzovnikov,[2] G. J. Ackland,[2] C. W. Monk,[1] and S. Anzellini[3]

AFFILIATIONS

[1]Materials and Physics Research Group, School of Science, Engineering and Environment, University of Salford, Manchester M5 4WT, United Kingdom
[2]SUPA, School of Physics & Astronomy and Centre for Science at Extreme Conditions, The University of Edinburgh, Edinburgh EH9 3FD, United Kingdom
[3]Diamond Light Source Ltd., Harwell Science and Innovation Campus, Diamond House, Didcot OX11 0DE, United Kingdom

[a]Author to whom correspondence should be addressed: j.e.proctor@salford.ac.uk

ABSTRACT

A comparison is made between the three principal methods for the analysis of neutron and x-ray diffraction data from noble gas fluids by direct Fourier transform. All three methods (standard Fourier transform, Lorch modification, and Soper–Barney modification) are used to analyze four different sets of diffraction data from noble gas fluids. The results are compared to the findings of a full-scale real-space structure determination, namely, Empirical Potential Structure Refinement. Conclusions are drawn on the relative merits of the three Fourier transform methods, what information can be reliably obtained using each method, and which method is most suitable for the analysis of different kinds of diffraction data. The mathematical validity of the Lorch method is critically analyzed.

I. INTRODUCTION

A. Analysis of fluid diffraction data via Fourier transform: The Q_{max} cutoff problem

On a fundamental level, the interpretation of diffraction data from fluids, network glasses, and amorphous solids is simple. The coherent scattering intensity $I_{coh}(Q)$ (of x rays or neutrons) is predicted by the Debye scattering equation, given below for a fluid composed of spherically symmetric identical particles:

$$I_{coh}(Q) = \sum_{m=1}^{N} \left[\sum_{n=1}^{N} \frac{f(Q)^2 \sin(Qr_{mn})}{Qr_{mn}} \right]. \quad (1)$$

Here, $f(Q)$ is the atomic form factor (x rays) or scattering length (neutrons) (tabulated data are available for this[1]) and N is the number of particles in the beam. The parameter r_{mn} refers to the distance between particles m and n. The structure factor $S(Q)$ is

obtained from the coherent scattering intensity $I_{coh}(Q)$ according to Eq. (2),[2,3]

$$S(Q) = \frac{I_{coh}(Q)}{Nf(Q)^2}. \quad (2)$$

If the coherent scattering intensity $I_{coh}(Q)$ is defined according to the Debye scattering equation, then the definition of $S(Q)$ according to Eq. (1) will result in $\lim_{Q \to \infty} S(Q) = 1$.

In reality $I_{coh}(Q)$, as defined by the Debye scattering equation, cannot be obtained. Instead, the experimentally measured scattering data undergo a series of corrections to calculate the raw coherent scattering intensity from the sample $I_{raw}(Q)$ as reliably as possible. The remaining normalization to obtain $S(Q)$ is performed using Eq. (3) through division by $N'f(Q)^2$, where N' is an arbitrary scaling parameter to ensure that $\lim_{Q \to \infty} S(Q) = 1$. The normalization via N' accounts for several effects including the fact that $I_{raw}(Q)$

13 October 2023 04:06:23

[unlike $I_{coh}(Q)$] depends on the incident beam intensity,

$$S(Q) = \frac{I_{raw}(Q)}{N'f(Q)^2}. \tag{3}$$

All further analysis utilizes this definition of $S(Q)$.

The radial distribution function $g(r)$ enables parameters such as the co-ordination number (CN) describing sample properties in real space to be obtained. The $g(r)$ function can be derived by a Fourier transform (FT) of the structure factor $S(Q)$ provided by the diffraction data. For a fluid composed of spherically symmetric identical particles, the FT relationship given in Eq. (4)[2,3] is exactly correct. The parameter g_0 is an arbitrary constant, which is eliminated in any case when $g(r)$ is normalized (as outlined in the supplementary material),

$$4\pi r^2 \left[g(r) - g_0 \right] = \frac{2}{\pi} \int_0^\infty Qr \left[S(Q) - 1 \right] \sin(Qr)dQ. \tag{4}$$

By definition, $g(r)$ represents the probability of finding particles separated by a particular distance. Therefore, $g(r)$ has certain *mathematical* properties, which most arbitrary functions do not. Notably, only functions which can be derived from an ensemble of point positions in space are possible $g(r)$ functions.

For fluids, there are further *physical* constraints: In the absence of long range order, one must have $g(r) = 1$ at long range; furthermore, $g(r)$ and its derivatives must be continuous. From Eq. (4), these constraints also apply to $S(Q)$.

We can, therefore, identify four distinct constraints on any meaningful $S(Q)$,

1. The associated $g(r)$ must correspond to an ensemble of points in space.
2. For atoms with finite size, $g(r)$ must be positive everywhere and go to zero for r approximately equal to the atomic diameter.
3. $S(Q)$ must be continuous.
4. Derivatives of $S(Q)$ must be continuous.

In fact, constraints 3 and 4 are consequences of constraint 1, but we list them separately because methods exist to resolve them. Only a very small fraction of possible functions satisfy these necessary constraints, and there is no reason to suppose that $S(Q)$ derived from real data via Eq. (3) will do so.

One problem is that real diffraction data cover a range in Q only from a finite minimum value of Q, Q_{min}, up to some finite value, Q_{max}. In this work, we will focus primarily on solutions to the problems caused by $Q_{max} < \infty$. We will, however, return to a discussion of the potential problems caused by $Q_{min} > 0$ in the conclusions. The Q_{max} problem is particularly severe for x-ray diffraction, where destructive interference between x rays scattered from different parts of the same atom prevents scattering at large Q. This phenomenon is represented mathematically using the parameter known as the atomic form factor $f(Q)$. Even for neutron diffraction, the range in Q covered by the data is limited by its relation to the scattering angle (which cannot exceed 180°) and by the reduction in $I_{coh}(Q)$ at large Q that takes place independently of

the effect of changes to $f(Q)$, via the $1/Q$ factor in the Debye scattering equation [Eq. (1)]. This phenomenon, in both x-ray and neutron diffraction experiments, is known as the Q_{max}-cutoff problem. This problem leads to all four types of unphysicalities described above.

In the absence of data, Eq. (4) can only be applied to experimental data if some theoretical assumption is made about $S(Q > Q_{max})$. The simplest theory is to take $[S(Q > Q_{max}) - 1] = 0$ (equivalently, integrate only to Q_{max}). This "absence of evidence is evidence of absence" theory is so widely used that it is often not even recognized as an assumption. It almost certainly violates all the physical constraints described above: most obviously it introduces a discontinuity in S(Q).

Mathematically, this sharp cutoff at Q_{max} means that the FT generates a "$g(r)$," which is a convolution of the true $g(r)$ data with the FT of a step function $f_S(Q)$: this latter is a sinc function which produces spurious oscillations at a frequency determined by Q_{max}. These manifest as peaks in $g(r)$ at unfeasibly small r, and in extreme cases, even negative $g(r)$ (both leading to $g(r)$ violating condition 2 above).

We can write the FT of data up to Q_{max} as

$$4\pi r^2 \left[g(r) - g_0 \right] = \frac{2}{\pi} \int_0^\infty Qr[S(Q) - 1]\sin(Qr)f_S(Q)dQ$$
$$= \frac{2}{\pi} \int_0^{Q_{max}} Qr[S(Q) - 1]\sin(Qr)f_S(Q)dQ, \tag{5}$$

where $f_S(Q)$ is the step function (see Fig. 1), which means that beyond Q_{max}, the integrand is zero and can be ignored. We note that if the data are extensive enough that $[S(Q > Q_{max}) - 1] = 0$,
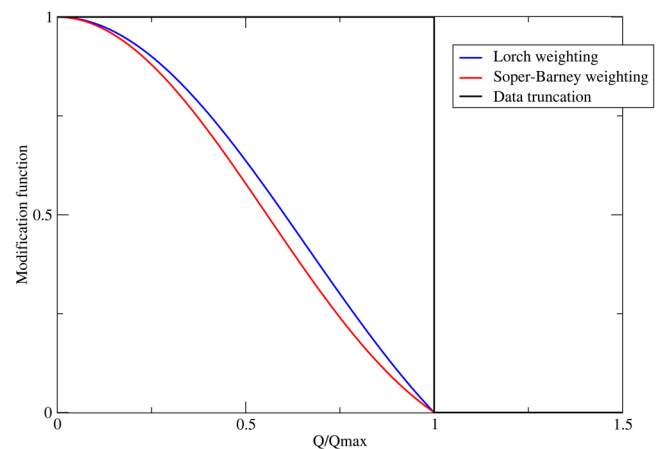
FIG. 1. Modification functions which permit the application of Eq. (4) in the absence of data for $Q > Q_{max}$, plotted as a function of Q/Q_{max}. The black line indicates a step function f_S which is discontinuous at Q_{max}, a method referred to as "direct FT." The blue line is the Lorch function f_L, and the red line is the Soper–Barney function f_{SB}. These are discontinuous in the first derivative. All functions continue to infinity with a value of zero.

this also gives a zero integrand above Q_{max} and the step function is unnecessary.

This discussion may appear unnecessarily pedantic, but its purpose is to introduce the concept of $f_S(Q)$. While the step function preserves the integrity of the data up to Q_{max}, *any* function that is equal to zero above Q_{max} makes it possible to perform a complete FT in spite of the missing data. As described, the step function (sinc convolution) may produce nonsensical results, which are typically undesirable. A different function, which modifies the data but does not introduce a discontinuity, may produce a $g(r)$ which is less obviously wrong.

B. Solution to the Q_{max} cutoff problem by the Lorch modification function

In 1969, Lorch proposed a solution[4] to the discontinuity arising from the Q_{max} cutoff problem, beginning from Eq. (4). Where we use $S(Q)$, Lorch used the notation $i(Q)$. Lorch's method is essentially to make a compromise. Instead of seeking to determine the exact value of $g(r)$ for all r, we will obtain the integrated value of $g(r)$ over a range of values about some central value r_0: $(r_0 - \Delta) \leq r \leq (r_0 + \Delta)$ (Lorch used $\Delta/2$ instead). In this case, both sides of Eq. (4) are integrated with respect to r within these limits and Lorch obtained an equation which is reproduced below using our notation as Eq. (6),

$$4\pi r_0^2 \left[g(r_0) - g_0 \right] = \frac{2}{\pi} \int_0^{Q_{max}} Qr_0[S(Q) - 1]\sin(Qr_0)\left[\frac{\sin(Q\Delta)}{Q\Delta}\right]dQ$$
$$f_L(Q\Delta) = \frac{\sin(Q\Delta)}{Q\Delta}. \tag{6}$$

The result of this compromise is an expression on the right hand side identical to Eq. (5) except that the integrand is now multiplied by a modification function $f_L(Q\Delta)$, the sinc function. $Q_{max} = \pi/\Delta$ is usually chosen as this makes $Q_{max}\Delta$ correspond to the first zero in the sinc function. In this case, constraint 3 as listed above will be satisfied: We now integrate $[S(Q) - 1]f_L$ which is continuous at Q_{max} instead of $[S(Q) - 1]f_S$, which is not. Beyond Q_{max}, the integrand of the FT must be zero: this can be achieved by some assumption on the data: $[S(Q > Q_{max}) - 1] = 0$, or by setting $f_L(Q > Q_{max}) = 0$. These yield identical results: to maintain compatibility with the step function approach, we adopt the second interpretation throughout.

Equation (6) has been used as a standard method to address the Q_{max}-cutoff discontinuity problem for the 50 years since Lorch's paper was published; it has 571 citations to date, 125 of which are from the last 5 years. The method has been used in many papers fundamental to our understanding of fluids, network glasses, and amorphous solids.

However, in 2011[5] it was shown that if the integral is performed as instructed in Lorch's original paper one in fact obtains the equation reproduced below in our notation as Eq. (7). Here, the application of a cutoff at some finite value of Q cannot be justified for the second integral, which is not necessarily small compared to the first integral.

$$4\pi r_0^2 \left[g(r_0) - g_0 \right]$$
$$= \frac{2}{\pi} \int_0^{Q_{max}} Qr_0[S(Q) - 1]\sin(Qr_0)\left[\frac{\sin(Q\Delta)}{Q\Delta}\right]dQ$$
$$+ \frac{2}{\pi} \int_0^{\infty} [S(Q) - 1]\left[\frac{\sin(Q\Delta)\cos(Qr_0)}{Q\Delta} - \cos(Qr_0)\cos(Q\Delta)\right]dQ. \tag{7}$$

Over the decades, Lorch's method has been very successful at removing unphysical oscillations in $g(r)$ functions but this does not prove that the method is producing $g(r)$ functions which are correct. This naturally raises the question: can Lorch's result [Eq. (6)] be recovered through the use of reasonable approximations? Let us examine the problem carefully.

To begin, we integrate Eq. (3) exactly as described within the specified limits [Eq. (8)]. The left hand side gives us the integrated value of $g(r)$ straightforwardly but, in order to continue referring to the non-integrated value of $g(r)$ [as is done in Ref. 4 and in Eqs. (6) and (7)], it is necessary to make additional assumptions. We assume that $\Delta \ll r_0$ and that $g(r)$ does not vary significantly over the region covered by the integral over r. Equation (9) specifies the exact, and approximate, relationships between $g(r)$ and its integrated equivalent which we shall label $G(r_0, \Delta)$,

$$\int_{r_0-\Delta}^{r_0+\Delta} 4\pi r_0^2 \left[g(r) - g_0 \right] dr = \frac{2}{\pi} \iint_{r=r_0-\Delta,Q=0}^{r=r_0+\Delta,Q=\infty} Qr[S(Q) - 1]\sin(Qr)dQdr, \tag{8}$$

$$G(r_0, \Delta) = \int_{r_0-\Delta}^{r_0+\Delta} 4\pi r_0^2[g(r) - g_0]dr,$$
$$G(r_0, \Delta) \approx 2\Delta \times 4\pi r_0^2[g(r_0) - g_0]. \tag{9}$$

The right hand side of Eq. (8) is a standard integral of the form $\int x\sin x dx$. Application of trigonometric identities $\cos(A + B) = \cos A \cos B - \sin A \sin B$, etc. results in Eq. (7), in agreement with Ref. 5.

Nevertheless, we will now see that the Lorch function is recoverable. If we implement the condition that $\Delta \ll r_0$ on the right hand side of Eq. (8) also, then the variation of the r term within the integrand is negligible over the range of the integral so we can replace it with r_0 and simply integrate $\sin(Qr)$ with respect to r. In this case, again by making use of standard trigonometric identities, Eq. (4) is recovered, i.e., the Lorch function is mathematically valid under this condition.

It is not correct to also set $\sin(Qr) \rightarrow \sin(Qr_0)$ in Eq. (8), as it would necessitate the far more stringent condition on Δ that $\Delta \rightarrow 0$. This is because we would require the variation in the value of Qr over the course of $r_0 - \Delta \leq r \leq r_0 + \Delta$ to cause negligible change to the value of $\sin(Qr)$ even in the high Q limit.

13 October 2023 04:06:23

C. Other solutions to the discontinuity at Q_{max} cutoff problem

Since the work of Lorch, several other solutions have been proposed to the Q_{max} cutoff problem. The most notable one is the modification function proposed by Soper and Barney.[5] Similarly to the Lorch function, this function involves accepting an averaged value of $g(r)$ over a range $\pm\Delta$. This function [Eq. (10)] takes values in the low Q and high Q limits identical to the sinc function proposed by Lorch [Eq. (6)] and is inserted into the FT equation in an analogous manner [Eq. (11)],

$$f_{SB}(Q\Delta) = \frac{3}{(Q\Delta)^3}(\sin(Q\Delta) - Q\Delta\cos(Q\Delta)), \quad (10)$$

$$4\pi r_0^2\left[g(r_0) - g_0\right] = \frac{2}{\pi}\int_0^{Q_{max}} Qr_0[S(Q) - 1]\sin(Qr_0)f_{SB}(Q\Delta)dQ. \quad (11)$$

Figure 1 shows the Lorch and Soper–Barney modification functions. Compared to the Lorch function, the Soper–Barney function has a wider central maximum (to second order in $Q\Delta$, $f_L = 1 - \frac{(Q\Delta)^2}{3!}$ and $f_{SB} = 1 - \frac{(Q\Delta)^2}{10}$) but then attenuates the data at higher Q far more than the Lorch function. If Δ is set analogously to the procedure used above with the Lorch function, then the wider central maximum causes a given value of Q_{max} to result in a larger value for Δ. Setting $f_{SB}(Q_{max}) = 0$ is essential to eliminate the discontinuity in $S(Q)$. This results in $\Delta = 4.49(1)/Q_{max}$, compared to $\Delta = \pi/Q_{max}$ for the Lorch method. In some cases, the Soper–Barney function has been implemented with the width Δ being r-dependent, though we will not pursue that here.

An alternative to the use of a modification function is to perform the FT [Eq. (5) or equivalent] to obtain $g(r)$, then perform the inverse FT to return to $S(Q)$, followed by an iterative process to transform back and forth seeking physically reasonable behavior [in particular, ensuring that $g(r) = 0$ within the atomic radius]. This approach was pioneered first by Kaplow et al.,[6] then outlined in more detail by Eggert et al.[2] The adjustable fitting parameters in this process are g_0 (representing the density) and two scaling factors specific to their method for background subtraction.

Unfortunately, to perform this analysis it was necessary to smooth the original $S(Q)$ data using a cubic-spline smoothing routine in which the amount of smoothing applied varied as a function of Q. In addition, the optimum values of the density and scaling factors varied substantially according to the value chosen for Q_{max}, not converging to a stable value (let alone the correct value for the density) within the range of Q_{max} studied (6–10 Å^{-1}). Due to these difficulties (in particular, the need for Q-dependent smoothing), we have not pursued the iterative approach in the present work. Iterative approaches are, however, available in the software packages Amorpheus[7] and LiquidDiffract.[8]

D. A Bayesian view on the modification functions

The Lorch method tackled the discontinuity problems (3 and 4 above) by modifying the measured $S(Q)$ by a function which goes to zero at Q_{max}. This has the cost that the data have to be adulterated to alleviate the problems of the discontinuity.

Another way to write the exact same transformation is

$$4\pi r^2\left[g(r) - g_0\right]$$
$$= \frac{2}{\pi}\int_0^\infty [Qr[S(Q) - 1]\sin(Qr)f_B(Q/Q_{max})$$
$$+ Qr[S_p(Q) - 1]\sin(Qr)[1 - f_B(Q/Q_{max})]]dQ, \quad (12)$$

with $S_p(Q) = 1$. The second term is introduced to connect the modification approach to a Bayesian framework. In the Bayesian approach to data, one starts with a prior assumption $[S_p(Q) - 1]$ for the data, and this is modified by the data. As well as the prior, an essential ingredient for the Bayesian method is the "strength" of the prior.

Taking f_B as the step function and the prior as "zero signal," we can read this equation as placing full weight on the data where it exists, up to Q_{max}, then full weight on the prior beyond Q_{max}. The Lorch and Soper–Barney modification can be seen as placing some weight on the prior at all Q., e.g., at $Q = Q_{max}/2$, these methods determine $g(r)$ by placing approximately equal weight to the prior as the real data.

Of course, there is no requirement for the prior to be chosen as zero, one might, e.g., choose the results of a simulation, or a previous experiment. Whatever prior is chosen, where one has full confidence in the data one can set $f_B = 1$ and the prior will be ignored, if one has no confidence in the data, e.g., because it does not exist, the non-zero prior means that the second term can provide non-zero contribution to the FT from $Q > Q_{max}$.

There are several approaches to the use of the prior: it can be used to combine theory and experiment with weights determined by the researcher, or it can be deployed in the spirit of Lorch/ Soper–Barney to remove the discontinuity.

In this second case, one might choose f_B to be the step function, and the prior to be a function which matches the data at Q_{max} and decays gracefully to zero (the form of the prior at $Q < Q_{max}$ is irrelevant, because it will be multiplied by zero). Such an approach retains the advantages of Lorch and Soper–Barney in addressing the Q_{max} cutoff problem, but does not compromise the integrity of the measured data.

E. The effect of noise in the $S(Q)$ data

Even if data are available up to extremely high Q, encompassing all oscillations in $S(Q)$, there remains the problem of noise. From the Debye scattering equation [Eq. (1)], we expect that the physically significant oscillations in $S(Q)$ will decay in amplitude upon Q increase. However, the noise amplitude in real experimental data does not decay on Q increase. In Eq. (4), $S(Q)$ is multiplied by Q in order to construct a valid FT to invert. The effect of this at high Q is simply to amplify noise. As we see throughout this work, it is common practice to smooth $S(Q)$ data so this problem can be avoided—including implementation of Q-dependent smoothing. This, however, raises a separate set of uncertainties: How can one be certain that features of physical significance have not also been lost in the smoothing process?

13 October 2023 04:06:23

F. Binning requirements for numerical integration to obtain g(r)

Regardless of whether we employ the standard FT method, Lorch modification function or Soper–Barney modification function, $g(r_0)$ is obtained by a numerical integration to $Q = Q_{max}$ of a function including $\sin(Qr_0)$. We, therefore, require the $S(Q)$ data to be binned at sufficiently fine resolution in Q such that a single cycle in the $\sin(Qr_0)$ function covers a large number of Q-values in the experimental data. Thus, if the binning interval is δQ we require

$$r_0 \delta Q \ll \pi. \tag{13}$$

Thus, we can see that the finite binning interval in the $S(Q)$ data can cause the numerical integration procedure to fail when used at large r_0.

G. Calculation of co-ordination number (CN)

The co-ordination number (CN) can be obtained from diffraction data via $g(r)$ using Eq. (14),

$$CN = 4\pi\rho \int_0^{r_{max}} r^2 g(r) dr. \tag{14}$$

Here, r_{max} is the radius of the minimum following the first and highest peak in $g(r)$. However, as shown in Eq. (14), this method can only be used if the density ρ is known reasonably accurately.

II. METHODS

We analyzed four different sets of neutron/x-ray diffraction $S(Q)$ data (described later). The numerical integrations to perform the FT and CN calculations were performed using our Octave code, which is available (with documentation) in the supplementary material. When using the Lorch and Soper–Barney modification functions $g(r)$ was obtained from $G(r, \Delta)$ via the approximate relationship in Eq. (9). In all cases, $g(r)$ was calculated for 1000 values of r, up to a maximum of 20 Å. Equations (5), (6), and (11) fail if we attempt to calculate $g(r = 0)$ so the minimum value of r was therefore 0.02 Å. Normalization was performed automatically by the code, simply by addition/subtraction to ensure $g(0.02 \text{ Å}) = 0$ then division by $g(20 \text{ Å})$ to ensure $g(r) = 1$ in the high r limit. In cases where ripples in $g(r)$ in the low r region caused $g(0.02 \text{ Å}) > g(20 \text{ Å})$ normalization and calculation of the CN were not attempted.

The CN was obtained by numerical integration of the first peak in $g(r)$. The integration limits for the calculation of the CN were obtained automatically by the code by locating the maximum value of $g(r)$, then the adjacent minimum values and integrating between these limits. This was chosen instead of integrating from $r = 0$ to reduce the effect of unphysical ripples in $g(r)$ at low r.

In all cases (see Sec. I D), the finite binning in the $S(Q)$ data could cause the $g(r)$ obtained at large r to be unreliable. When FT

has been performed using the Lorch or Soper–Barney modification function, $g(r)$ may be unreliable at low r (we require $r \gg \Delta$). Allowing ourselves a factor of 10 in these criteria, we obtain the conditions given in Eq. (15), which are marked on all graphs of $g(r)$ where they are within the range covered in the plot,

$$
\begin{aligned}
r_{max} &= \frac{\pi}{10\delta Q}, \\
r_{min} &= \frac{10\pi}{Q_{max}},
\end{aligned}
\tag{15}
$$

The datasets were as follows:

- **Dataset A** is a neutron $S(Q)$ from liquid Ar at 85 K, ambient pressure (a density of 2.13×10^{28} at./m^3) from Ref. 9. The data are collected to 12 Å^{-1}, covering all features in $S(Q)$. The binning is $\delta Q = 0.0294$ Å^{-1}, leading to $r_{max} = 10.7(1)$ Å. The data were smoothed by the authors of Ref. 9 using a method communicated to them via private communication, and not described in Ref. 9.
- **Dataset B** is a set of synchrotron x-ray diffraction $S(Q)$ data from fluid Ar at 300 K covering eight pressures from 46 to 830 MPa (densities from 9.89×10^{27} to 2.59×10^{28} at./m^3). The binning is $\delta Q = 2.1 \times 10^{-3}$ Å^{-1}, leading to $r_{max} = 149.6(1)$ Å. Since these data have not been published previously, we outline here how $S(Q)$ was obtained from the raw data. The raw coherent scattering intensity $I_{raw}(Q)$ was obtained from the experimental data by subtraction of the background signal from either the empty DAC or the DAC containing solid Ar following masking of the solid Ar Bragg peaks. The structure factor $S(Q)$ was obtained by obtaining $I_{raw}(Q)/f(Q)^2$ and normalizing [i.e., accounting for the factor of N' in Eq. (3)] to ensure appropriate behavior in the high-Q limit. We have three comments to make on this process:

1. The $f(Q)$ values are obtained from tabulated data in the International Tables for Crystallography.[1] To obtain $f(Q)$ at the exact values of Q corresponding to our experimental data, it is necessary to find an empirical equation that fits the tabulated data, or to do a linear interpolation between the datapoints. We used a linear interpolation. The easier alternative of fitting an empirical equation has some pitfalls. The tabulated $f(Q)$ data in the literature are far more closely spaced at low Q, so it would be easy to overfit to these data at the expense of a good fit to the more sparse data at high Q. In addition, since $f(Q) \rightarrow 0$ at high Q, a small error in $f(Q)$ in absolute terms will be large as a proportion of $f(Q)$ and lead to noise in $I_{raw}(Q)$ being amplified by an arbitrary amount.

2. Since the signal-to-noise ratio becomes very poor due to the decreasing $f(Q)$ by the second peak in $I_{raw}(Q)$, it is challenging to accurately normalize $S(Q)$ in the high Q limit. We normalized by measuring the peak scattering intensity from the second peak in $I_{raw}(Q)$ and the minimum scattering intensity from the trough following this peak and normalizing to make the average of these values equal to 1. Clearly, this normalization choice is arbitrary, however, in the absence of S(Q) data encompassing all oscillations in S(Q) there is no possible

normalization method that does not involve some arbitrary choice.

3. At the lowest values of Q at which data are present, $I_{raw}(Q) < 0$ in most cases, due to difficulties with the background subtraction procedure. A small constant [1%–5% of the peak value of $S(Q)$] was therefore added to set $S(Q) \geq 0$ for all Q.

Further information on dataset B, including figures, and experimental details, are given in the supplementary material.

- **Dataset C** is a set of neutron diffraction $S(Q)$ data from supercritical fluid Kr[10] covering 17 pressures up to 20 MPa and densities up to 6.2×10^{27} at./m^3 at 300 K. The data are collected to $Q = 4$ Å^{-1}. We have analyzed selected data from this set using Empirical Potential Structure Refinement (EPSR)[11,12] and published this elsewhere.[13] The binning is $\delta Q = 0.05$ Å^{-1}, leading to $r_{max} = 6.3(1)$ Å. The data were smoothed by visual observation by the original authors prior to presentation in tabulated form in Ref. 10.

- **Dataset D** is a set of neutron diffraction $S(Q)$ data from supercritical fluid Kr[13] collected up to significantly higher pressure and density at 310 K (seven pressures ranging from 40 up to 200 MPa, 1.62×10^{28} at./m^3). We have published an EPSR analysis of all these data elsewhere.[13] The binning is $\delta Q = 0.05$ Å^{-1}, leading to $r_{max} = 6.3(1)$ Å. In contrast to dataset C, these data have not been smoothed.

In each case, the pressures have been measured experimentally and the densities were calculated from the pressures using the relevant fundamental equation of state via NIST REFPROP[14] (Ref. 15 for Ar and Ref. 16 for Kr). In the case of dataset C, the densities given in the source (Ref. 10) were calculated from the experimentally measured pressures using an earlier equation of state.[17] We have calculated the measured pressures from the densities using Ref. 17 and then recalculated the densities using the fundamental EOS.[16] This has resulted in a change of ca. 10% in the higher calculated densities.

The pressure–temperature phase diagrams of Ar (from Ref. 18) and Kr (compiled for this work using the methodology presented in Ref. 18) are given in the supplementary material, with the P,T points marked at which the datasets in the present study were collected.

III. RESULTS

A. Dataset A

To begin, we performed a standard FT of the entirety of dataset A. Figure 2 shows the $g(r)$ obtained from this FT, with the original $S(Q)$ shown in the inset. The FT result is in agreement with the $g(r)$ shown in Ref. 9 (the source of dataset A). The $g(r)$ obtained is also in agreement with that obtained using real-space structure determination implemented in the Dissolve package[19] (see the supplementary material) and leads to a CN of 11.9. This is clearly a reasonable value for dataset A; however, it is worth noting that even with this very good quality data there are unphysical oscillations present in $g(r)$ for $r < 3$ Å, where $g(r) = 0$ is expected. On one level, this does not matter (we do not need a FT to tell us
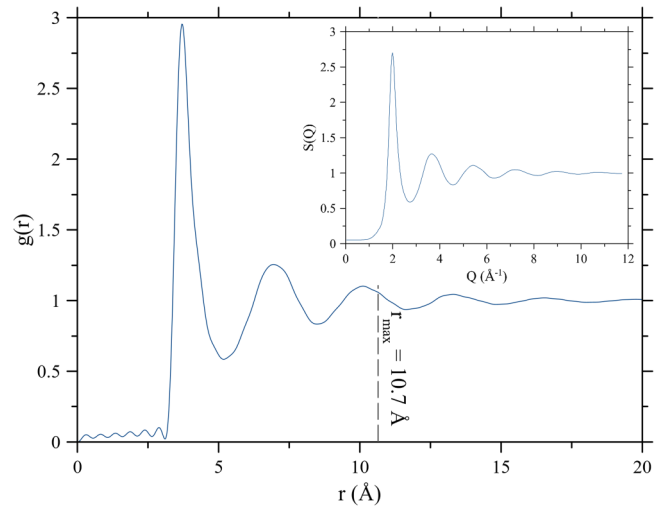

FIG. 2. Normalized $g(r)$ for liquid Argon at 85 K (dataset A) obtained by direct FT (i.e., assuming that $[S(Q > Q_{max}) - 1] = 0$) of the full neutron $S(Q)$ (the inset).

that atoms are not allowed to overlap) but it does introduce an error in the normalization of $g(r)$, and hence CN calculation.

Attempting the standard FT using cutoffs at successively lower Q leads to the ripples at $r < 3$ Å becoming larger, until at a Q cutoff of 8 Å^{-1} the obtained $g(r)$ is no longer physically realistic, due to the low-r ripples preventing accurate normalization. Figure 3 shows the $g(r)$ obtained with the standard FT, Lorch, and Soper–Barney modifications with Dataset A at 8 Å^{-1} Q-cutoff. The Lorch and Soper–Barney modification functions both give
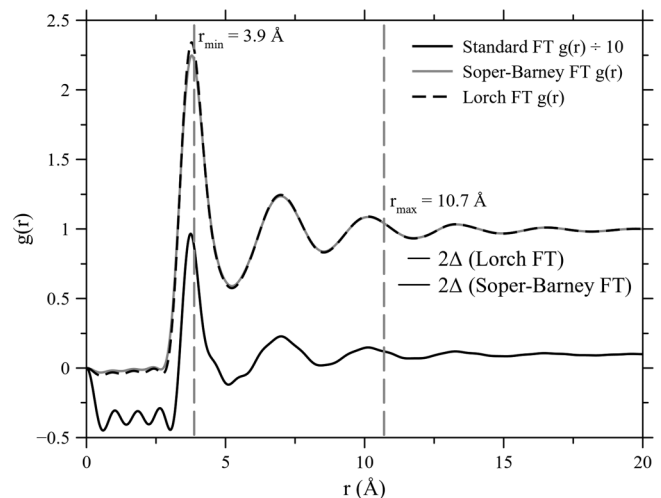

FIG. 3. Normalized $g(r)$ for liquid Argon at 85 K (dataset A) obtained from the neutron $S(Q)$ truncated at 8 Å^{-1}. FT with abrupt truncation Lorch and Soper–Barney modification are shown. The standard FT $g(r)$ is ÷10 for clarity.

reasonable results at cutoffs for which the standard FT fails. The functions give very similar results, the only significant difference being that the principal maximum in $g(r)$ is slightly wider for the Soper–Barney modification function due to the $8\,\text{Å}^{-1}$ Q-cutoff resulting in a slightly larger value for Δ.

Our final investigation with this dataset is to study how low the Q_{max} cutoff can be made with each modification function while still obtaining a physically reasonable $g(r)$. Using both the Lorch and Soper–Barney modification functions, even a cutoff as low as $3\,\text{Å}^{-1}$ does not result in ripples at low r as bad as those shown for the standard FT in Fig. 3. A more rigorous test is whether the CN calculated remains reasonable. Figure 4 shows the CNs calculated using all three methods for Q_{max} cutoffs all the way from 3 to $12\,\text{Å}^{-1}$. At $12\,\text{Å}^{-1}$ (as previously discussed), all ripples in $S(Q)$ are included in the data and the pure FT gives a $g(r)$ in agreement with that resulting from EPSR analysis of the data. This leads to a CN of 11.9—as would be expected for a liquid near the triple point. So it is reasonable to take this as the correct value. At $12\,\text{Å}^{-1}$ cutoff, both modification functions produce a CN slightly larger than the maximum physically realistic value of 12; however, as the cutoff is reduced to $5\,\text{Å}^{-1}$ the $g(r)$ at least remains stable at this value around 2.5% too large when the FT is done with a modification function. In contrast, the CN oscillates wildly when the FT is performed using the standard method.

B. Dataset B

Remaining with fluid Ar, dataset B is a set of synchrotron x-ray data (unlike datasets A, C, and D which are all neutron data) collected at nine pressures up to $830\,\text{MPa}$. Due to the rapid decrease in scattering intensity upon Q increase (caused primarily

by the decrease in the atomic form factor $f(Q)$), the signal-to-noise ratio beyond the first peak in $I_{raw}(Q)$ is very poor. Only the highest pressure $S(Q)$ produced a g(r) that could be normalized using the method outlined earlier. Even in this case, to achieve a g(r) that could be normalized it was necessary to adopt the Lorch or Soper–Barney modification function and to use a Q_{max} cutoff that only included the first peak in S(Q).

Figure 5(a) shows the normalized $S(Q)$ functions at 830 and 444 MPa, illustrating the difficulties caused by the need to normalize on the basis of noisy data at high Q. The first peak in S(Q) is more intense at the lower pressure.

To obtain any reasonable $g(r)$ function, it was necessary to use the S(Q) data only up to a Q_{max} at the minimum after the first
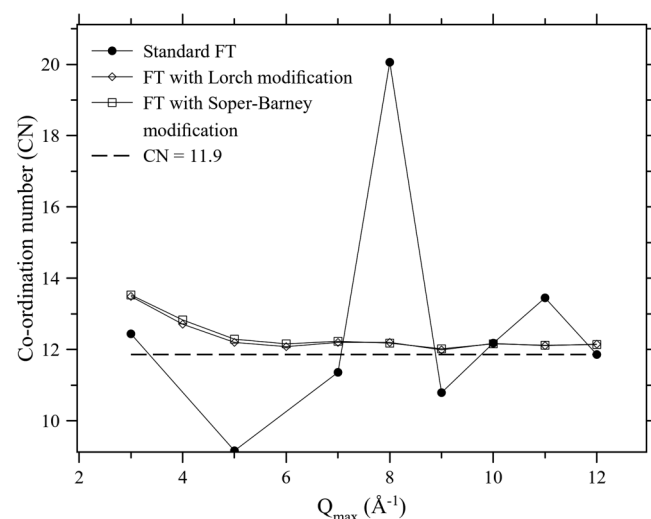
FIG. 4. CN of liquid Ar at 85 K (dataset A) obtained by standard FT of the neutron diffraction data, and FT utilizing the Lorch and Soper–Barney modification functions for a variety of Q_{max} cutoffs. The $g(r)$ functions produced by the standard FT for 4 and $5\,\text{Å}^{-1}$ cutoff could not be normalized so calculation of the CN was not attempted.
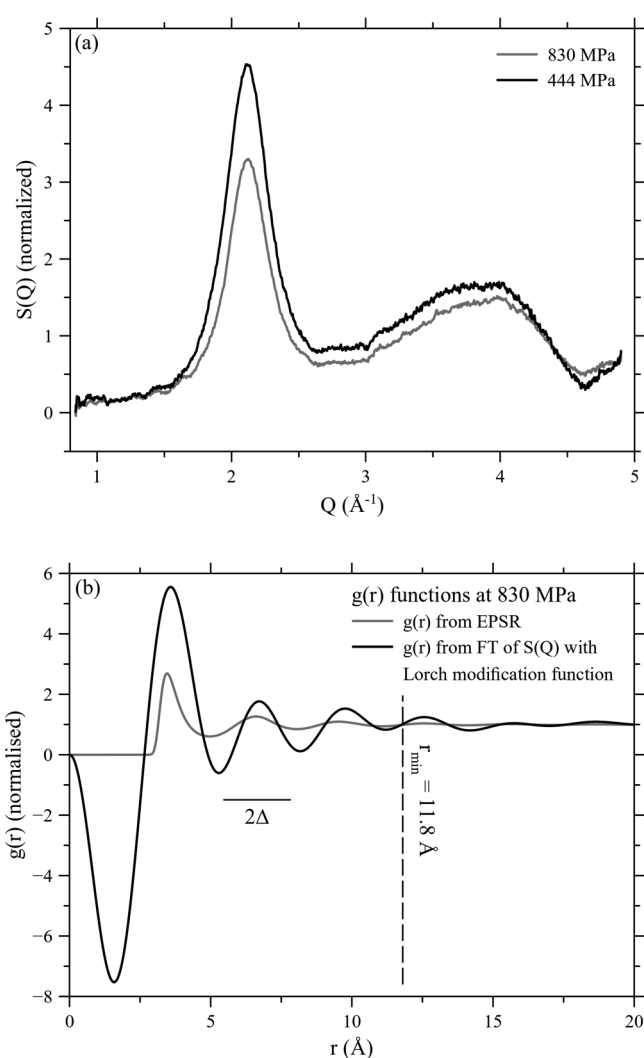
FIG. 5. (a) Normalized $S(Q)$ from dataset B at 830 and 444 MPa. (b) $g(r)$ functions at 830 MPa produced from FT of the $S(Q)$ with Lorch modification (cutoff of $2.66\,\text{Å}^{-1}$) and from EPSR using $I_{raw}(Q)$.
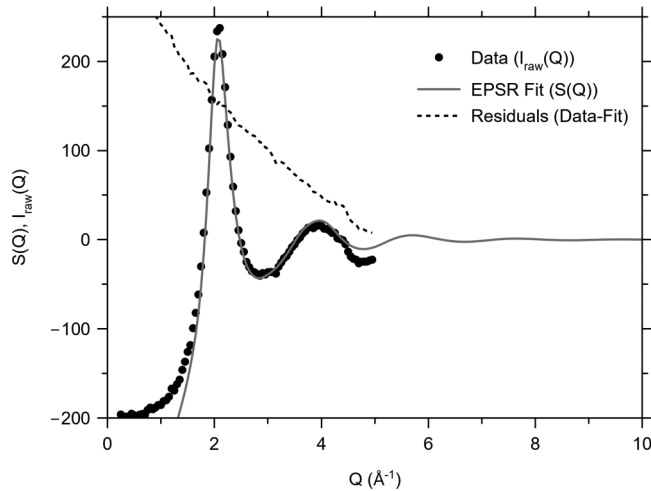
FIG. 6. The experimental $I_{raw}(Q)$ data and the fitted $S(Q)$ from EPSR at 444 MPa, demonstrating the excellent quality of fit achieved by EPSR when given the measured $I_{raw}(Q)$, and residuals highlighting the lack of corrections.

peak in $S(Q)$ (i.e., ca. 2.5–3 Å^{-1}). For this reason, analysis could only be performed using the Lorch and Soper–Barney modification functions. Even with this methodology, only the highest pressure yielded a $g(r)$ that could be normalized. The obtained CN varied massively depending on which modification function was used: 25.4 with the Lorch modification function and 19.7 with the Soper–Barney modification function. Both of these are far larger than is physically realistic.

Figure 5(b) shows the normalized g(r) obtained using FT with the Soper–Barney modification function at 830 MPa. Although the normalization procedure is mathematically valid, it is clear that due to the size of the ripple at low r and the existence of negative g(r) in some regions the resulting function is not physically realistic.

We have also performed a full EPSR refinement of the datasets at 830 and 444 MPa, with densities of 43 049 and 37 341 mol/m^3. Due to the aforementioned issues with the normalization of the $S(Q)$, we have decided to entirely bypass this and provide to EPSR directly the $I_{raw}(Q)$ as measured. In principle, any lack of correction applied to the $I_{raw}(Q)$ should come out of EPSR as a meaningful trend in the residuals. We employed a 5000 Ar atom simulation box and used the OPLS-Noble Gases forcefield as the reference potential (the same potential parameters as included in Dissolve). Once the box equilibrated, empirical potential (EP) refinement was allowed until a fit that was deemed satisfactory was achieved. At this point, the EP was frozen and prevented from further changing, and the structure was sampled over 5000 accumulations. We achieved an excellent quality of fit to the data, yielding a physically reasonable $g(r)$ and an associated CN of 11.42(15) at 444 MPa. This is below 12, not exceeding the maximum justifiable coordination and is what we would expect on the liquid-like side of the Frenkel line. As expected, the residuals from the fitting process show a clear functional Q-dependence, highlighting the lack of appropriate corrections that should have been performed to obtain

$S(Q)$ from $I_{raw}(Q)$. An example $g(r)$ resulting from the EPSR is shown in Fig. 5(b) and an example S(Q) (with residuals) is shown in Fig. 6.

C. Dataset C

Datasets C and D are both from fluid Kr. Dataset C is collected to a relatively low value of Q (4 Å^{-1}), at which significant oscillations in $S(Q)$ are still present. It is, therefore, not surprising that the standard FT produces a physically unrealistic result, compared to performing the transform using the Lorch method [Fig. 7 shows the transforms for the highest density point in this dataset, and the $S(Q)$ data]. The standard FT produces massive oscillations in the low r limit, and also smaller oscillations in the high r limit which are not physically realistic given the gas-like density (less than half the density of the Frenkel line in Kr at 300 K). Due to the fact that the $S(Q)$ data extend only to 4 Å^{-1}, combined with the wide binning interval, the low and high-r regions in which $g(r)$ may be unreliable due to binning and Q-cutoff issues, respectively, overlap for these data.

Figure 8 shows the CNs obtained by Fourier transform of Dataset C at densities from 2.8 to 6.1×10^{27} at./m^3. This corresponds to pressures up to ca. 20 MPa, not reaching the liquid-like side of the Widom lines, let alone the Frenkel line. We would, therefore, expect the CN to vary in a smooth and monotonic manner throughout. This is the case when the Lorch or Soper–Barney modification functions are employed. Data are also available in Ref. 10 at four lower densities; however, none of the FT procedures utilized here produced physically realistic $g(r)$ functions in these cases. The standard FT resulted in a $g(r)$ that could not be normalized, while the Lorch and Soper–Barney modification functions produced a $g(r)$ with no clear minimum following the first peak. This resulted in subsequent peaks also being included in the
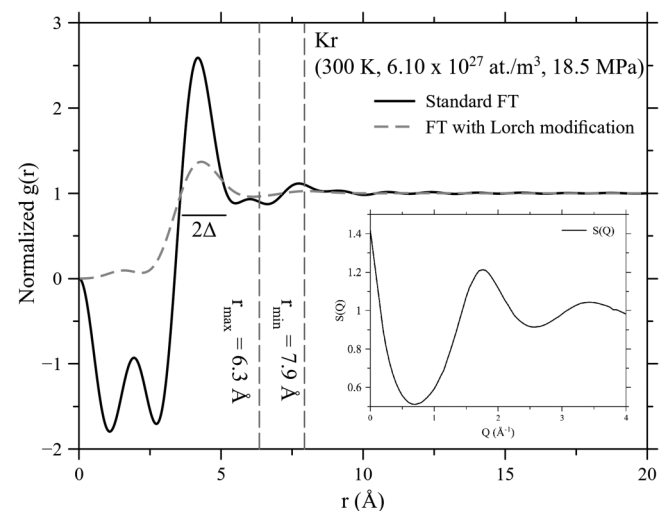

FIG. 7. Normalized $g(r)$ obtained from direct FT and FT with Lorch modification of neutron $S(Q)$ from supercritical fluid Kr at 300 K (6.10×10^{27} at./m^3 density).[10] Inset: $S(Q)$ data from Ref. 10.
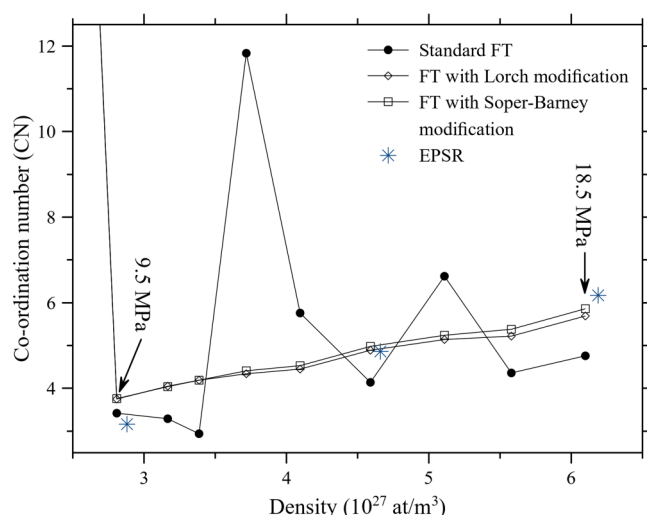
FIG. 8. CNs obtained by Fourier transform (present work) and EPSR (Ref. 13) of dataset C (neutron diffraction data from Kr at 300 K to 4 Å^{-1} from Ref. 10).

integration to obtain the CN and an unrealistically high value being obtained.

We have published an analysis of selected neutron diffraction data from Ref. 10 elsewhere.[13] The CNs obtained from EPSR are in reasonable agreement with those obtained by direct FT via both Lorch and Soper–Barney modification functions (Fig. 8). It is hard to discern the extent to which quantitative agreement exists between the $g(r)$ functions produced by modified FT and EPSR due to the $\pm\Delta$ broadening in r caused by the modification process (Fig. 7).
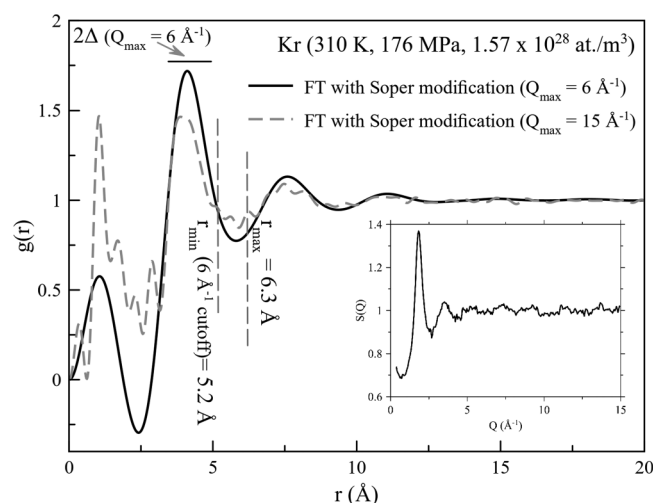

FIG. 9. $g(r)$ functions for fluid Kr at 310 K, 176 MPa obtained by FT with Soper modification function with different cutoffs as indicated. Inset: $S(Q)$ data for Kr at 310 K, 176 MPa.

D. Dataset D

Finally, we present our analysis of dataset D. The large amount of noise in this data prevents meaningful results being obtained by direct Fourier transform, despite the large Q_{max} (20 Å^{-1}). Beginning with a high density datapoint [176 MPa $(1.57 \times 10^{28}$ at./m$^3)$] we can see that above 6 Å^{-1} the data are just noise [Fig. 9, the inset shows the $S(Q)$ data to 15 Å^{-1}]. In Fig. 9, we show $g(r)$ functions obtained with the Soper–Barney modification function for two cutoffs (6 and 15 Å^{-1}). The data for 15 Å^{-1} are ruined by ripples resulting from Fourier transforming noise, while the data for 6 Å^{-1} exhibit a principal peak in $g(r)$ that is somewhat larger than the ripples and leads to a CN of 12.4 (our EPSR analysis of this data published previously indicated a CN of 11.7). The highest density datapoint (200 MPa) can be analyzed in a similar manner (CN of 12.5) but for the next lowest pressure (148 MPa) the Fourier transform produces a $g(r)$ that cannot be normalized.

IV. CONCLUSIONS

The results and analysis presented here throw light on some important, but hitherto neglected, aspects of the analysis of fluid x-ray and neutron diffraction data by FT from reciprocal space to real space. In the analysis of x-ray data, and even (to some extent) neutron data, various arbitrary choices are unavoidably made in the analysis. It is, therefore, difficult to judge the reliability or otherwise of $g(r)$ functions generated by FT of diffraction data without knowledge of the details of these arbitrary choices: In particular, the methodology used to normalize $S(Q)$ and $g(r)$, and the methodology used to obtain $f(Q)$ at the required Q-values from the tabulated data in the literature. In addition, it is essential to know the resolution in the binning of the experimental data as this determines the validity of the resulting $g(r)$ function at high r.

Moving beyond these issues, the analysis of diffraction data by direct FT, or by FT using a modification function, is always subject to some extent to the Q_{max}-cutoff problem. We find many cases where the Lorch and Soper–Barney modification functions can produce a non-pathological $g(r)$ while direct FT cannot and despite concerns raised in recent years, the Lorch modification function is a mathematically rigorous method to analyze fluid diffraction data.

However, mathematical rigor is no guarantee of physical correctness. We have identified four ways in which the Q_{max}-cutoff problem can produce physically unreasonable $g(r)$ functions when diffraction data from disordered materials are analyzed by FT, but only one of these [the continuity of $S(Q)$ at Q_{max}] is addressed by the Lorch and Soper–Barney modification functions. Neither method attempts to address the issue of missing data beyond Q_{max}, the discontinuity in slope of $S(Q)$. Neither method guarantees that the $g(r)$ corresponds to any possible arrangement of atoms in space, certainly not the correct one. Neither method guarantees positive $g(r)$, nor zero $g(r)$ at unphysically small distances. We found that the $g(r)$ functions produced by these methods are highly sensitive to choice of Q_{max}, and strong evidence that the results have not converged for typical values of Q_{max}.

The fundamental problem with purely data-driven Fourier methods is that they necessarily require some theoretical assumption about the missing data above Q_{max}. All the methods used here assume that $[S(Q > Q_{max}) - 1] = 0$ and seek only to address the

discontinuity. These mathematical tricks remove some pathological errors in $g(r)$, but the sensitivity to Q_{max} gives little confidence that the non-pathological $g(r)$ are accurate.

The mathematical elegance of these models should not disguise the fact that they serve to suppress the high-Q data. It can be argued that measurement of $[S(Q) - 1]$ may be less reliable at high-Q; however, the solution to unreliable data is to downweight its importance in analysis, not to pretend that it is smaller than measured, or even zero, which is what the Lorch and Soper–Barney methods do. Indeed, by "modifying" the data away from the experiment in a way designed to avoid pathologies, rather than based on physics, an inaccuracy (broadening) in $g(r)$ is caused.

The problem of missing or unreliable data can be approached with Bayesian-type methods. These provide a framework to combine experimental and theoretical results with appropriate Q-dependent weights. They also indicate a way to remove the Q_{max} discontinuities without modifying the data.

Nevertheless, direct FT methods can address only two of the four constraints on $g(r)$ which must be satisfied. To our knowledge, there is no method to determine whether a given function "$g(r)$" or "$S(Q)$" can correspond to any 3D ensemble of atomic arrangements. Therefore, assuming one wishes the data analysis to produce a $g(r)$ which respects the atomic theory of matter, one needs to use a method beginning from real-space configurations. EPSR[11,12] is an example of this approach. It exploits the fact that for any given interatomic potential, $V(r)$, both $g(r)$ and $S(Q)$ are uniquely defined and respect the atomic theory of matter. So EPSR seeks to find a $V(r)$ consistent with $S(Q)$, and a $g(r)$ which automatically satisfies all four unphysicality issues. EPSR has elements of the Bayesian approach, with the prior being a potential that is modified by the data to produce a best fit to the measured $S(Q)$ from functions constrained to be physically reasonable.

Comparison of EPSR results for all four datasets has shown that EPSR can provide physically realistic results even for very poor quality data for which analysis by FT fails. It appears that, historically, $S(Q)$ data have often been smoothed to allow analysis by FT. Results presented here indicate that EPSR can provide physically realistic results with the unsmoothed data. Therefore, if smoothing of the $S(Q)$ data is necessary to allow analysis by FT then EPSR should be attempted instead.

Finally, while the Lorch and Soper–Barney modification functions make some progress toward addressing the Q_{max}-cutoff problem, neither address the other cutoff problem: The fact that real $S(Q)$ data do not extend to $Q = 0$. In any real diffraction experiment, $I_{raw}(Q)$ data cannot be collected at $Q \approx 0$ as there is no way to distinguish between x rays/neutrons that have been transmitted without interacting with the sample, and those scattered at low Q. While $\lim_{Q \to 0} S(Q) \neq 0$, the contribution to the value of $[g(r) - g_0]$ from low-Q scattering in Eq. (4) (and modified FT functions derived from it) *does* vanish due to the presence of the $\sin(Qr)$ term in the integral so the substitution of the zero lower limit in the integral with a small but finite minimum value of Q should be acceptable. However, it is worth noting that analysis methods beginning from real-space configurations and predicting $S(Q)$ in the Q-range for which data exists also avoid this potential pitfall.

In future, the work presented here can be extended to cover other fluids comprised of spherically symmetric particles such as metallic fluids and CH_4, and molecular fluids for which different partial $S(Q)$ data are obtained by isotopic substitution in neutron diffraction.

SUPPLEMENTARY MATERIAL

Supplementary material is provided, consisting of $g(r)$ function obtained by performing EPSR on dataset A, full description of the experimental methods and analysis to obtain $S(Q)$ for dataset B, the Octave/Matlab code used, and documentation for this code.

AUTHOR DECLARATIONS

Conflicts of Interest

The authors have no conflicts to disclose.

Author Contributions

J.E.P. designed the study, performed experiments, wrote code, performed data analysis by FT, and wrote the manuscript. C.G.P. designed the study, performed experiments, performed data analysis by EPSR and contributed to the manuscript. B.M. wrote code and performed data analysis by FT. M.A.K. performed data analysis and prepared figures. G.J.A. designed the study, contributed to the manuscript and prepared figures. C.W.M. prepared figures and SA performed experiments. All authors approved the final version of the manuscript.

J. E. Proctor: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (supporting); Supervision (lead); Writing – original draft (lead); Writing – review & editing (equal). **C. G. Pruteanu:** Conceptualization (supporting); Formal analysis (equal); Methodology (supporting); Writing – review & editing (equal). **B. Moss:** Formal analysis (supporting); Software (lead). **M. A. Kuzovnikov:** Formal analysis (equal); Writing – original draft (supporting). **G. J. Ackland:** Conceptualization (supporting); Funding acquisition (lead); Methodology (supporting); Writing – original draft (supporting); Writing – review & editing (equal). **C. W. Monk:** Data curation (supporting); Writing – original draft (supporting). **S. Anzellini:** Data curation (supporting); Investigation (supporting).

DATA AVAILABILITY

The data that support the findings of the study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. J. C. Wilson and E. Price, *International Tables for Crystallography* (Kluwer, Dordrecht/International Union of Crystallography, 1999), Vol. C.

[2] J. H. Eggert, G. Weck, P. Loubeyre, and M. Mezouar, Phys. Rev. B **65**, 174105 (2002).

[3] B. E. Warren, *X-ray Diffraction* (Dover Publications, 1990).

[4] E. Lorch, J. Phys. C: Solid State Phys. **2**, 229 (1969).

[5] A. K. Soper and E. R. Barney, J. Appl. Crystallogr. **44**, 714 (2011).

[6] R. Kaplow, S. L. Strong, and B. L. Averbach, Phys. Rev. **138**, A1336 (1965).

[7] S. Boccato *et al.*, High Pressure Res. **42**, 69 (2022).

[8] B. J. Heinen and J. W. E. Drewitt, Phys. Chem. Miner. **49**, 9 (2022).

[9] J. L. Yarnell, M. J. Katz, R. G. Wenzel, and S. H. Koenig, Phys. Rev. A **7**, 2130 (1973).

[10] A. Teitsma and P. A. Egelstaff, Phys. Rev. A **21**, 367 (1980).

[11] A. K. Soper, Chem. Phys. **202**, 295 (1996).

[12] A. K. Soper, Mol. Simul. **38**, 1171 (2012).

[13] C. G. Pruteanu, J. S. Loveday, G. J. Ackland, and J. E. Proctor, J. Phys. Chem. Lett. **13**, 8284 (2022).

[14] E. Lemmon, I. H. Bell, M. Huber, and M. McLinden, NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version10.0, Standard Reference Data Program, National Institute of Standards and Technology: Gaithersburg, 2018.

[15] C. Tegeler, R. Span, and W. Wagner, J. Phys. Chem. Ref. Data **28**, 779 (1999).

[16] E. W. Lemmon and R. Span, J. Chem. Eng. Data **51**, 785 (2006).

[17] N. J. Trappeniers, T. Wassenaar, and G. J. Wolkers, Physica **32**, 1503 (1966).

[18] J. E. Proctor, *The Liquid and Supercritical Fluid States of Matter* (CRC Press, 2020).

[19] T. Youngs, Mol. Phys. **117**, 3464 (2019).

13 October 2023 04:06:23