Nyquist-Shannon sampling theorem applied to refinements of the atomic pair distribution function

Christopher L. Farrow,[1] Margaret Shaw,[2] Hyunjeong Kim,[3] Pavol Juhás,[1] and Simon J. L. Billinge[1,4,*]

[1]*Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA*

[2]*Department of Physics and Astronomy, Princeton University, Princeton, New Jersey 08544, USA*

[3]*Energy Technology Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8565, Japan*

[4]*Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA*

(Received 30 May 2011; published 18 October 2011)

We have systematically studied the optimal real-space sampling of atomic pair distribution (PDF) data by comparing refinement results from oversampled and resampled data. Based on nickel and a complex perovskite system, we show that not only is the optimal sampling bounded by the Nyquist interval described by the Nyquist-Shannon (NS) sampling theorem as expected, but near this sampling interval, the data points in the PDF are minimally correlated, which results in more reliable uncertainty estimates in the modeling. Surprisingly, we find that PDF refinements quickly become unstable for data on coarser grids. Although the Nyquist-Shannon sampling theorem is well known, it has not been applied to PDF refinements, despite the growing popularity of the PDF method and its adoption in a growing number of communities. Here, we give explicit expressions for the application of NS sampling theorem to the PDF case, and establish through modeling that it is working in practice, which lays the groundwork for this to become more widely adopted. This has implications for the speed and complexity of possible refinements that can be carried out many times faster than currently with no loss of information, and it establishes a theoretically sound limit on the amount of information contained in the PDF that will prevent over-parametrization during modeling.

DOI: 10.1103/PhysRevB.84.134105 PACS number(s): 61.05.cp, 02.30.Nw, 61.05.fm

I. INTRODUCTION

Atomic pair distribution function (PDF) analysis of x-ray and neutron powder diffraction data is becoming prominent in structure analysis of complex materials due to an increasing interest in studying nanoscale structural order.[1] Details of the atomic arrangement play a crucial role in determining the physical properties of materials and for crystalline materials, we have powerful crystallographic methods for solving the structure with high precision.[2,3] Equivalent methods are lacking for nanomaterials such as nanoparticles and nanoporous materials,[1] which limits our ability to optimize and fully exploit their interesting properties. Furthermore, it is becoming increasingly apparent that nanoscale fluctuations exist in many bulk materials and that these also are important factors in the properties of those materials.[4–7] The PDF technique has emerged as a powerful tool for extracting quantitative information from these materials when high-performance, modern sources of high-energy x rays and neutrons are coupled with emerging data modeling software.[8–10] For example, details of nanoparticle structure, defects, size, and strain state can be quantitatively extracted from the smallest nanoparticles,[11] nanosized domains can be studied in melt-quenched pharmaceutical drugs,[12] species intercalated into nanoporous hosts can be characterized,[13,14] and hidden symmetry breaking nanodomains can be detected in technologically important bulk materials.[6,7] As the community of users grows, dedicated experimental facilities are appearing for PDF studies[15,16] as well as specialized software.[17–22] As the PDF is becoming a recognized tool for structure characterization in a growing number of scientific communities,[7,12,23–25] it is important to reevaluate and strengthen our analysis techniques.

The PDF is a sine Fourier transform of properly corrected and normalized x-ray or neutron powder diffraction data. One of the user-specified parameters in the Fourier transform step is the grid of points on which the PDF is calculated. Currently, the sampling grid for PDFs is typically chosen in an *ad hoc* way, for example, to give a visually smooth PDF. The information content in the PDF does not increase for grid intervals above a critical value given by the Nyquist-Shannon (NS) sampling theorem. If the data are oversampled, not only is no new information introduced, the points in the PDF become statistically correlated,[26,27] which leads to improper estimates of uncertainties in refinement parameters as well as slowing down structural refinements.[28] Despite this, in practice, the NS theorem is rarely, if ever, taken into account in PDF modeling. There is no examination in the literature of how it applies explicitly to the PDF. Such a treatment for the EXAFS case[29] has had a large impact on how modeling is practiced in that field and we hope that this treatment will similarly influence practices in the growing PDF community.

We have systematically studied through modeling the optimal PDF sampling interval for PDF data and demonstrate that it is consistent with the value predicted by the NS sampling theorem.[30] This gives the minimum amount of information we need to completely specify a PDF from a given measurable scattering function. When this optimal sampling is enforced, we see significant speedup in our PDF refinements accompanied by a small increase in estimated uncertainties due to the reduction of statistical correlations among the PDF points. When the data are made sparser than the optimal sampling interval, the refinement results rapidly become unreliable due to aliasing.

II. THE PDF METHOD

The PDF method is a total scattering technique for determining local order in nanostructured materials.[27] The technique does not require periodicity, so it is well suited for studying nanoscale features in a variety of materials.[8,9] The experimental PDF, denoted $G(r)$, is the truncated Fourier

transform of the total scattering structure function, $F(Q) = Q[S(Q) - 1]$:[31]

$$G(r) = \frac{2}{\pi} \int_{Q_{\min}}^{Q_{\max}} F(Q) \sin(Qr)\, dQ, \tag{1}$$

where Q is the magnitude of the scattering momentum. The structure function $S(Q)$ is extracted from the Bragg and diffuse components of x-ray, neutron, or electron powder diffraction intensity. For elastic scattering, $Q = 4\pi \sin(\theta)/\lambda$, where λ is the scattering wavelength and 2θ is the scattering angle. In practice, values of $Q_{\min}$ and $Q_{\max}$ are determined by the experimental setup and $Q_{\max}$ is often reduced below the experimental maximum to eliminate noisy data from the PDF, since the signal-to-noise ratio becomes unfavorable in the high-Q region.

The PDF gives the scaled probability of finding two atoms in a material a distance r apart and is related to the density of atom pairs in the material.[27] For a macroscopic scatterer, $G(r)$ can be calculated from a known structure model according to

$$\begin{aligned} G(r) &= 4\pi r\, [\rho(r) - \rho_0], \\ \rho(r) &= \frac{1}{4\pi r^2 N} \sum_i \sum_{j \neq i} \frac{b_i b_j}{\langle b \rangle^2} \delta(r - r_{ij}). \end{aligned} \tag{2}$$

Here, ρ_0 is the atomic number density of the material and $\rho(r)$ is the atomic pair density, which is the mean weighted density of neighbor atoms at distance r from an atom at the origin. The sums in $\rho(r)$ run over all atoms in the sample, b_i is the scattering factor of atom i, $\langle b \rangle$ is the average scattering factor, and r_{ij} is the distance between atoms i and j.

In practice, we use Eq. (2) to fit the PDF generated from a structure model to a PDF determined from experiment. For this purpose, the delta functions in Eq. (2) are Gaussian-broadened and the equation is modified to account for experimental effects. PDF modeling is performed by adjusting the parameters of the structure model, such as the lattice constants, atom positions, and anisotropic atomic displacement parameters to maximize the agreement between the theoretical and an experimental PDF. This procedure is implemented in PDFGUI,[17] which is the program used in this study. PDFGUI uses the Levenberg-Marquardt algorithm[32,33] to locally optimize the model structure. The algorithm also provides estimates of uncertainties on those parameters upon convergence, though, strictly, the estimates are only accurate if the data are independent and the statistical errors are Gaussian distributed and properly determined.[28]

III. THE NYQUIST-SHANNON SAMPLING THEOREM

The Nyquist-Shannon sampling theorem specifies an upper bound on the sampling interval of a discretized signal in the time domain such that the sample contains all the available frequency information from the signal. This upper bound is $\pi/\Delta\omega$, where $\Delta\omega$ is the angular frequency bandwidth of the signal.[30] The quantity $\pi/\Delta\omega$ is commonly referred to as the Nyquist interval. A continuous or discrete signal sampled on a grid finer than the Nyquist interval can be, in principle, perfectly reconstructed via interpolation, since the sampling does not compromise the information content of the signal.

In relation to the PDF, the angular frequency domain is Q space and we are interested in sampling in r space, the analog of the time domain. The frequency information is specified by $F(Q)$ [see Eq. (1)], which has bandwidth $Q_{\max}$. (The sampling theorem as presented in Shannon's paper deals with signals having positive and negative frequency components. The bandwidth is defined as the maximum absolute frequency value. Mathematically, $F(Q)$ is an odd function [see Eq. (15) in Ref. 31], a fact we use when transforming $F(Q)$ to $G(r)$ [Eq. (1)]. The "full" spectrum of $F(Q)$ that includes the negative-frequency branch can be calculated from the positive-frequency branch, and spans the range $[-Q_{\max}, Q_{\max}]$. $Q_{\min}$ does not enter into this, since we enforce $F(Q < Q_{\min}) = 0$ during modeling.[31]) This gives a Nyquist interval of

$$dr_N = \pi/Q_{\max}. \tag{3}$$

The sampling theorem states that the PDF can be sampled on any grid with intervals shorter than this without losing any information from $F(Q)$.

Whittaker[34] and Shannon[30] describe an interpolation formula for reconstructing a signal from samples taken on a grid with interval, dr, less than the Nyquist interval. In terms of the PDF, the reconstruction formula is

$$G'(r) = \sum_n G(ndr) \frac{\sin[\pi(r/dr - n)]}{\pi(r/dr - n)}, \tag{4}$$

where n iterates over the points of the sample. Later we will demonstrate the benefits of modeling the PDF on an optimally sampled grid. This formula allows us to interpolate a model PDF onto a denser grid, e.g., for convenient visual inspection. In practice, the sampled data must extend beyond the desired range to avoid reconstruction errors in the high-r region.

A. Aliasing

Sampling $G(r)$ at or coarser than the Nyquist interval results in aliasing. This term refers to how, in undersampled data, high Q information in $F(Q)$ can masquerade as intensity at lower Q. This is demonstrated for the PDF by considering its Fourier series over $-r_{\max} \leqslant r \leqslant r_{\max}$. We choose this range because it lets us consider the sine-Fourier series [$G(r)$ is odd] and because the PDF over this range contains the same information as the PDF over $0 \leqslant r \leqslant r_{\max}$. Now,

$$G(r) = \sum_{m=1}^{m_{\max}} b_m \sin(Q_m r),$$

where $Q_m = m\pi/r_{\max}$. Since $G(r)$ contains no frequency components greater than $Q_{\max}$, $Q_m \leqslant Q_{\max}$, and thus $m_{\max} \leqslant Q_{\max} r_{\max}/\pi$.

Consider the mth term of the series sampled on the interval $dr = \pi/Q'$, where Q' and m are chosen such that $Q' \leqslant Q_m \leqslant Q_{\max}$. For the nth sample, the contribution to the Fourier series is $b_m \sin(ndr\, Q_m)$. Given the relationship between Q_m and Q', $ndr\, Q_m \geqslant n(\pi/Q')Q' = n\pi$. Thus we can represent the argument as $n\pi + (Q_m - Q')\, ndr = 2n\pi + (Q_m - 2Q')\, ndr$, so that the mth frequency component of the sample looks like $-b_m \sin[(2Q' - Q_m)ndr]$ for all n. The contribution to $G(r)$ from $F(Q)$ at $Q = Q_m$ therefore appears in $G(r)$ as if it came from $Q = 2Q' - Q_m$ in $F(Q)$. In $F(Q)$,

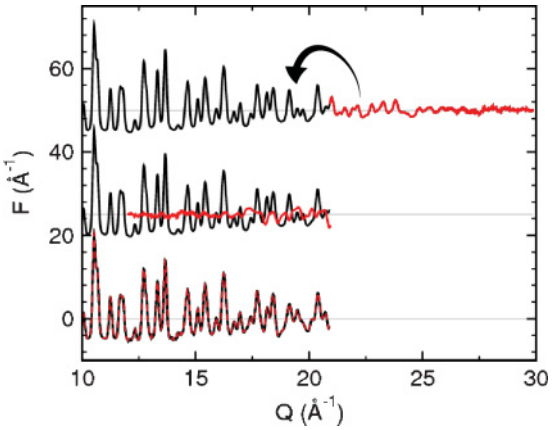

FIG. 1. (Color online) Demonstration of aliasing in $F(Q)$. (Top) Experimental nickel $F(Q)$ with $Q_{\max} = 29.9$ Å^{-1} featuring regions above and below $Q' = 20.9$ Å^{-1}. (Center) Experimental nickel $F(Q)$ with the region above Q' "folded" over to lower Q. (Bottom) Aliased $F(Q)$ obtained by sampling the PDF from the experimental $F(Q)$ on a grid with interval $dr = 0.15$ Å and Fourier transforming back to $F(Q)$ (solid line). This sampling interval is larger than the Nyquist interval ($dr_N = 0.105$ Å) and corresponds to $Q' = \pi/dr = 20.9$ Å^{-1}. Overlaid is the $F(Q)$ obtained by adding the unfolded and folded segments of the experimental $F(Q)$ (dashed line). Note that the Q axis starts at 10 Å^{-1}.

the signal above Q' gets "folded" back to lower Q and overlaps with the signal in the range $2Q' - Q_{\max} \leqslant Q \leqslant 2Q'$. This explains how information in $F(Q)$ is progressively lost in $G(r)$ if it is calculated on grids that are too coarse. The more undersampled the data the greater the Q range that is folded back and the greater the loss of information in $G(r)$ due to overlapping signals from different Q values. The effect is illustrated in Fig. 1.

We note that the case where the data are sampled precisely on a grid with the Nyquist interval, $dr = dr_N$, then $Q' = Q_m = Q_{\max}$ and there is no folding. However, there is still loss of information, since $\sin(Q_m n dr) = 0$, and so the mth Fourier amplitude b_m can take on any value. This is why a strict inequality between the sampling and the Nyquist intervals is required to avoid aliasing: $dr < dr_N$.

Aliasing implies that the sampled signal does not uniquely identify its source. Since some frequency components alias others, the PDF could represent the aliased $F(Q)$ just as well as the unaliased one. When back-Fourier transforming a sparsely sampled $G(r)$ into Q space, the aliased $F(Q)$ will result. The sampling theorem states that aliasing does not occur when sampling at an interval smaller than the Nyquist interval.

B. Structural information in the PDF

The sampling theorem determines the number of data points required to reconstruct a PDF signal from samples, which is

$$N = \Delta r/dr_N = \frac{\Delta r Q_{\max}}{\pi}, \tag{5}$$

where Δr is the extent of the PDF in r space. What is more relevant to PDF modeling is the amount of structural information in the PDF. N is an upper bound on this, since we cannot extract more independent observations of the structure than raw information from the signal. Given perfect data and the proper model, one can meaningfully extract N structural parameters from a PDF signal.

Factors such as noise and peak overlap can obscure the structural information in the PDF and therefore determine whether N is a good estimate of the amount of structural information in the PDF. For example, consider a situation where the PDF contains a single peak, but has a very large $Q_{\max}$. In this case, a complete structure model cannot be obtained from fitting this single peak, no matter how large N is. In another extreme case, imagine that the majority of PDF peaks have a single point or no points due to a small $Q_{\max}$. In this situation, the position and shape of the peaks cannot be determined with certainty.

In practice, the amount of structural information in the PDF cannot be precisely known. To perform a reliable refinement, the signal-to-noise ratio must be favorable,[27] the PDF peaks must be apparent, and the fit range must be such that the structural features one is seeking to model are accessible. In addition to this, we recommend using Rietveld refinement guidelines when refining the PDF, which advise that the ratio of independent observations to the number of refinement parameters should be around three to five, preferring the latter.[35]

IV. EXPERIMENTAL VERIFICATION

Powder diffraction data were collected from nickel (Ni) and $LaMnO_3$ (LMO) samples. The nickel data were collected using the rapid-acquisition pair distribution function (RaPDF) technique[36] with synchrotron x rays on beamline 6-ID-D at the Advanced Photon Source at Argonne National Laboratory. The sample was purchased from Alfa Aesar. The powdered sample was packed in a flat plate holder with thickness of 1.0 mm and sealed between Kapton tapes. Data were collected at room temperature in transmission geometry with an x-ray energy of 98.001 keV ($\lambda = 0.12651$ Å). An image plate camera (Mar345) with diameter of 345 mm was mounted orthogonally to the beam with a sample-to-detector distance of 178.4 mm.

The raw 2D data were reduced to 1D integrated intensity profiles using the FIT2D program.[37] Corrections for environmental scattering, incoherent and multiple scattering, polarization and absorption were performed according to the standard procedures[27] using PDFGETX2[18] to obtain the PDF with $Q_{\max} = 29.9$ Å^{-1}. This corresponds to $dr_N = 0.105$ Å.

The LMO data were collected using time-of-flight neutron diffraction at the NPDF instrument at the Los Alamos Neutron Scattering Center at Los Alamos National Laboratory. The LMO sample preparation and data collection have been described in detail elsewhere.[38] The LMO PDFs were produced with PDFGETN[20] using $Q_{\max} = 32.0$ Å^{-1}. This corresponds to $dr_N = 0.0982$ Å. Note that PDF sampling intervals coming from the NS theorem are around $dr_N = 0.1$ Å, which is ten times larger than the value of $dr = 0.01$ Å, which is default in PDFGETN[20] and PDFGETX2.[18]

In each case, experimental PDFs were generated with $r_{\max} = 20$ Å using $dr = 0.01$ Å. PDF data on sparser grids were created by removing points from this PDF in order to get the desired sampling interval. Pruning the data in this way is equivalent to recalculating the PDF from $F(Q)$ on the sparser

grid. We produced 31 data sets with varying dr against which models were refined.

We took as a reference data set the PDF generated on the default grid of $dr = 0.01$ Å and structural models were refined to the data. We then refined the same models to data sets on sparser grids. We define $\Delta_p(dr)$ for a parameter p as the absolute difference between the value of the parameter p refined for the data set sampled at interval dr and that refined for the reference data set. The accuracy of the refined parameters becomes unacceptable when $\Delta_p(dr)$ exceeds the statistical uncertainty on the difference, $\sigma[\Delta_p(dr)]$. This is given by $\sigma[\Delta_p(dr)] = \sqrt{\sigma^2[p(dr)] + \sigma^2[p(0.01)]}$, where $\sigma[p(dr)]$ and $\sigma[p(0.01)]$ are the estimated uncertainties on parameter p taken from the refinement for the data set sampled at interval dr and the reference data set, respectively. To determine if a refined parameter extracted from a sparse data set is accurate, we define a parameter quality factor $\mathcal{Q}_p(dr) = \Delta_p(dr)/\sigma[\Delta_p(dr)]$. If $\mathcal{Q}_p(dr)$ is less than or equal to one, the parameter value refined from the data set sampled at interval dr is within the expected uncertainty of the best estimate and is considered accurate. If $\mathcal{Q}_p(i)$ is greater than one, the change in the parameter's value is greater than the expected uncertainty, and the result is considered unreliable.

The parameter quality measure $\mathcal{Q}_p(i)$ is biased due to a couple of assumptions. First, by comparing all results with the refinement of the undiluted data, we assume that this refinement gives the best estimate for each parameter. The validity of this assumption is dependent on the systematic bias of the refinement results due to the quality of the data and the suitability of the refinement model. Since this bias is present in the diluted data as well, its effects should be negligible. Second, we assume that the uncertainty value derived from the refinement results is accurate. We discuss later that the uncertainty values derived from refinements of oversampled data sets are too small. This inflates the estimated quality factor when the data are oversampled, but does not invalidate the accompanying results.

The refinements from unaltered and sampled data sets were performed identically over a range from $r_{\min} = 0.01$ Å to $r_{\max} = 20.0$ Å using the program PDFGUI.[17] For the Ni data, the lattice parameter, isotropic atomic displacement parameter (ADP), dynamic correlation factor, scale factor, and resolution factor were varied in the refinements. In the LMO fits, three lattice parameters, four isotropic ADPs (one for each of the La, Mn, and axial and planar oxygen atoms), and seven fractional coordinates were varied along with the scale and correlation factors (see Ref. 39). From Eq. (5), we get that refinements over this range, $\Delta r = 19.99$ Å, yield $N_{\text{Ni}} = 191$ and $N_{\text{LMO}} = 203$. For the Ni data set, we have an observation-to-parameter ratio (OPR) greater than 30 and for LMO, the OPR is greater than 10. The refinements are therefore comfortably overconstrained and the optimization problem is well conditioned.

Various refinements were timed to measure the speedup in the program execution due to sampling.

V. RESULTS

When the Ni and LMO data are made sparser, the PDF profiles appear less smooth and the detailed shape of the peak profiles becomes less apparent. This is shown in Figs. 2 and 3. The data in panel (a) in both figures are on the reference grid ($dr = 0.01$ Å) and are both smooth and have well defined Gaussian-like peaks.[27] The data in panel (b) are sampled with $dr = 0.1$ Å, close to the Nyquist interval, and are not nearly as smooth, though the peaks are still well defined. Lastly, the data in panel (c) are sampled with $dr = 0.3$ Å, where there is apparent loss of information. The refined parameters from these fits are given in Tables I and II. Note that the uncertainty in the refined parameters increases from $dr = 0.01$ Å to $dr = 0.1$ Å, although each of these data sets produce acceptable results.

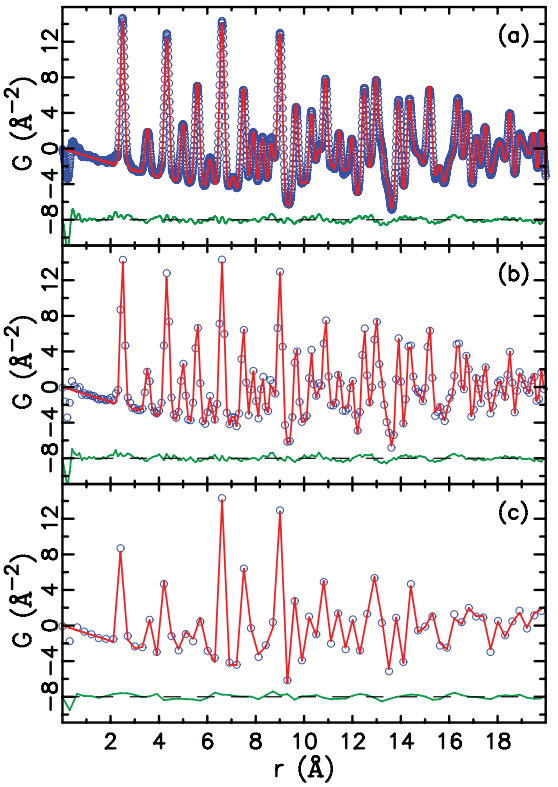

FIG. 2. (Color online) Fits to sampled Ni PDFs. (a) Unaltered data with $dr = 0.01$ Å. (b) Sampled data with $dr = 0.1$ Å. (c) Sampled data with $dr = 0.3$ Å. The data are shown as circles, the fits are the lines through the data, and the difference is shown offset below. All fits are of similar quality, despite the poor visual quality of the data in panels (b) and (c). The data shown in panel (c) are undersampled and produced unacceptably uncertain results, though this is not apparent from the difference curve.

In Fig. 4, we show the parameter quality values $\mathcal{Q}_p(i)$ plotted against the sampling interval. The quality factor is satisfactory for data sets that are sampled with grids close to the reference data set. This indicates that these refinements are producing the same parameter values. As the Nyquist interval is crossed (indicated in each case by the vertical dashed line), various quality factors rapidly become unacceptable. The figures dramatically show how well the NS theorem is obeyed. Identical refinements (within the uncertainties) are obtained on all sampling grids finer than dr_N, but the refinements rapidly degrade on coarser grids. The computation time of the refinements decreases rapidly with increasing sampling

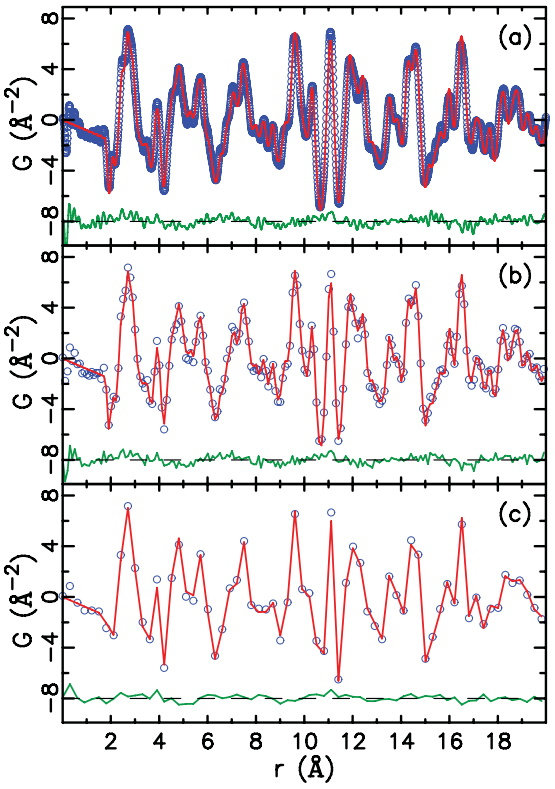

FIG. 3. (Color online) Fits to sampled $LaMnO_3$ PDFs. (a) Unaltered data with $dr = 0.01$ Å. (b) Sampled data with $dr = 0.1$ Å. (c) Sampled data with $dr = 0.3$ Å. The data are shown as circles, the fits are the lines through the data, and the difference is shown offset below. All fits are of similar quality, despite the poor visual quality of the data in panels (b) and (c). The data shown in panel (b) and (c) are undersampled, and the data in panel (c) produced unacceptably uncertain results. Note that in panel (c) several peaks are not resolved.

interval following a $1/dr$ law, as evident by the green line in the figure.

VI. DISCUSSION

Figure 4 verifies that the onset of unreliable refinements coincides with the Nyquist interval. The refined parameter values are all acceptable, and largely independent of the sampling interval in the oversampling region ($dr < dr_N$).

TABLE I. Parameters from Ni refinements using data with various dr. The Nyquist interval dr_N is 0.105 Å. Here, a denotes the lattice parameter, U_{iso} the isotropic ADP, δ_2 the vibrational correlation parameter, "scale" the data scale, and Q_{damp} the experimental resolution factor.

dr (Å)	0.01	0.10	0.12	0.30
R_w	0.112	0.120	0.119	0.084
a (Å)	3.53159(2)	3.53158(6)	3.53158(6)	3.53186(10)
U_{iso} (Å^2)	0.005446(7)	0.00545(2)	0.00543(2)	0.00570(4)
δ_2 (Å^2)	2.25(2)	2.20(5)	2.15(5)	2.2(2)
scale	0.7324(7)	0.733(2)	0.734(3)	0.761(4)
Q_{damp} (Å^{-1})	0.06307(11)	0.0632(4)	0.0634(4)	0.0653(7)

TABLE II. Parameters from $LaMnO_3$ refinements using data with various dr. The Nyquist interval, dr_N, is 0.0982 Å. Here, a, b, and c denote the lattice parameters, U_{iso} is the isotropic ADP (one for each primitive atom), x, y, and z are the fractional atomic coordinates, δ_2 is the vibrational correlation parameter, and "scale" the data scale.

dr (Å)	0.01	0.10	0.12	0.30
R_w	0.135	0.138	0.143	0.103
a (Å)	5.5394(2)	5.5394(6)	5.5393(7)	5.5362(14)
b (Å)	5.7441(2)	5.7443(7)	5.7442(8)	5.7536(13)
c (Å)	7.7059(2)	7.7059(9)	7.7054(10)	7.697(2)
δ_2 (Å^2)	2.44(3)	2.38(9)	2.35(9)	2.49(14)
scale	0.7941(11)	0.794(3)	0.795(4)	0.803(6)
La				
x	0.99234(10)	0.9923(3)	0.9926(4)	0.9917(6)
y	0.04828(8)	0.0482(2)	0.0481(3)	0.0469(5)
U_{iso} (Å^2)	0.00508(4)	0.00506(13)	0.0052(2)	0.0055(2)
Mn				
U_{iso} (Å^2)	0.00376(7)	0.0038(2)	0.0038(2)	0.0024(3)
O$_1$				
x	0.07300(11)	0.0730(4)	0.0730(4)	0.0739(7)
y	0.48625(10)	0.4862(3)	0.4864(4)	0.4874(7)
U_{iso} (Å^2)	0.00682(8)	0.0067(3)	0.0068(3)	0.0075(3)
O$_2$				
x	0.72515(8)	0.7251(2)	0.7252(3)	0.7247(5)
y	0.30682(8)	0.3068(3)	0.3069(3)	0.3072(5)
z	0.03876(6)	0.0388(2)	0.0389(2)	0.0399(3)
U_{iso} (Å^2)	0.00689(4)	0.0069(2)	0.0068(2)	0.0062(2)

Figures 2–4 indicate that visual appearance alone is not a good indicator of data quality. From Tables I and II, we see a decrease in R_w, the goodness of fit parameter, for the largest sampling interval. This apparent improvement in fit quality is a consequence of having fewer points to fit with the same number of fitting parameters. It is important to note that even at the extreme sampling interval of $dr = 0.3$ Å the refinements are evidently overconstrained, with an OPR near 11 for the Ni refinement and near four for the LMO refinement. These subtle contradictions emphasize the importance of observing the Shannon-Nyquist sampling theorem in PDF analysis.

The sampling theorem tells us that the information content in the data does not change as long as we sample on a grid finer than the Nyquist interval. We expect to and do refine the same parameters from such samples. As the data are sampled onto grids coarser than the Nyquist interval, we expect to lose structural information gradually. In contrast, refined values of the parameters become unreliable quickly as the Nyquist interval is exceeded. This is somewhat surprising, since the refinements are overconstrained even when sampled at three times the Nyquist interval. In Fig. 4, we see the quality of the refined parameters diverge well before this point. Intuition would tell us that it is possible to lose a considerable quantity of information by undersampling before refinements become unstable. This is not observed. The degradation of the refinements is not caused solely by information loss, but by information corruption due to aliasing and the current results show that this has a dramatic effect on the quality of the refined parameters.

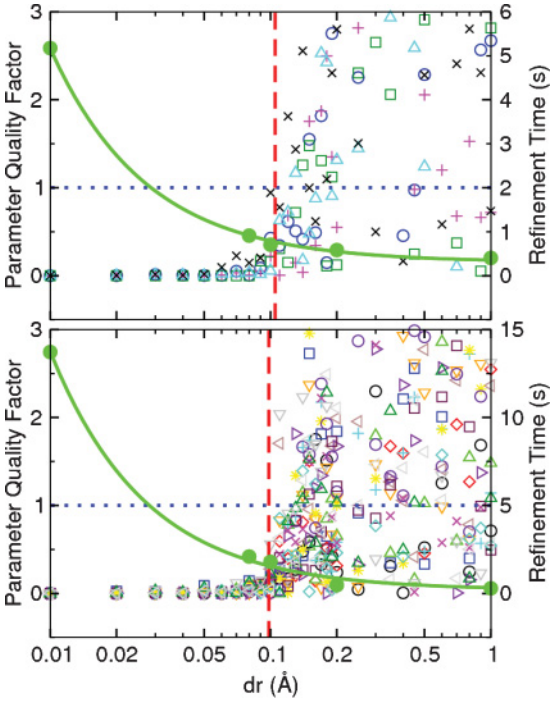

FIG. 4. (Color online) Refined parameter quality (open symbols) and refinement times (solid circles) measured using sampled Ni (top) and $LaMnO_3$ (bottom) data. The dotted horizontal line shows the cutoff between acceptable and unacceptable parameter quality. The dashed vertical line shows the value of dr_N predicted by the sampling theorem. For dr values larger than this, the quality of some parameters transition into the unacceptable region. The time values demonstrate the decrease in refinement time with increasing dr with more than a seven-fold speedup near dr_N. The solid curve through the time values is fit to the form $a + b/dr$.

Aliasing has two effects on a PDF signal, as described in Sec. III A. Foremost, aliasing lowers the effective maximum Q value in $F(Q)$ from $Q_{\max}$ to $Q' = \pi/dr$. This creates the obvious effect of lower resolution in the PDF, as seen in Figs. 2 and 3. In extreme cases, this will lead to poorly defined peaks in the PDF. Less obviously, sampling on a grid coarser than the Nyquist interval allows for the possibility that the PDF has originated from a different, aliased, $F(Q)$ as shown in Fig. 1. When calculating the model PDF, we enforce $F(Q > Q_{\max}) = 0$. When there is aliasing, the structure function resulting from $G(r)$ has $F(Q > \pi/dr) = 0$ and extra intensity below π/dr. Thus, aliasing results in finding biased parameters that describe the corrupted structural information. This is true regardless of the optimization algorithm used.

The estimated uncertainties on the fitting parameters for dr in the region of stable refinements are dependent on the sampling interval. We see from Tables I and II that the uncertainties on the parameters increase when estimated from the data sampled near the Nyquist interval compared to the reference data. The sampling theorem gives the number of data points necessary to fully represent the PDF. Any data sampled on a grid finer than the Nyquist interval are necessarily redundant. If a set of fitting parameters reproduces a particular set of points well on an optimal grid, those parameters will also reproduce the associated redundant points well. By not taking into account the statistical correlations between data points,[26] as in this study, this results in the fitting program underestimating uncertainty values on parameters. In principle, this can be overcome by propagating a full $N \times N$ variance-covariance matrix through the Fourier transform,[26] accounting for the statistical correlations between all points. This is computationally expensive and is not generally done. As the sampling interval increases (or equivalently as $Q_{\max}$ increases for a given sampling grid) the variance-covariance matrix becomes more banded around the diagonal, with significant correlations appearing only between points near to each other in the PDF. A larger sampling grid therefore reduces the statistical correlations between points in the PDF. Refining optimally sampled data at just below dr_N, therefore, not only results in correctly refined parameters, but also the best uncertainty estimates possible, in the absence of a full treatment of the covariances.

A fortunate side effect of refining optimally sampled data is a decreased refinement time. Shown in Fig. 4 is a plot of refinement times for some chosen sampling intervals. The trend in the plot shows that refinement time is proportional to the inverse of dr (shown as the broad solid line), or directly proportional to the number of data points, with a constant offset. This trend reflects the fact that the calculation of the PDF grows linearly with the number of sample points. Carrying out refinements on optimally sampled data gives a significant speed increase compared to the reference data; in this case, the speed increases by more than a factor of seven.

These observations indicate that PDF refinements should be performed on the sparsest grid possible with sampling interval less than the Nyquist interval. To produce an esthetically pleasing presentation of the PDF, one can always interpolate onto a finer grid using the Whittaker-Shannon interpolation formula [see Eq. (4)].

VII. CONCLUSIONS

The purpose of this research was to demonstrate the consequences of the Nyquist-Shannon sampling theorem as they apply to the PDF. We show that the quality of refined parameters diverges when sampling the PDF at intervals larger than the Nyquist interval, which is the result of aliasing. Furthermore, we show that the estimated uncertainties of refined parameters are more reliable when the PDF is optimally sampled. Statistically reliable uncertainties on refined parameters can be obtained by taking into account the correlations between all the points in $G(r)$,[26] but this comes at the computational expense of inverting a large error matrix. By optimally sampling the PDF, the correlations among points in the PDF are minimized, while preserving all the available structural information. This gives improved uncertainty estimates without costly computation, and may expedite refinements when the PDF can be computed over fewer points.

The Nyquist-Shannon sampling theorem gives an upper bound on the amount of structural information contained in an experimental PDF. This determines the Q and r extents that are required for a model refinement to be overconstrained. Oversampling the PDF does not add more information to a refinement, and therefore provides no benefit other than an esthetically pleasing visualization. This result emphasizes the

importance of collecting diffraction data to high Q when it is to be used for PDF modeling, since a larger Q_{max} decreases the Nyquist interval and makes accessible more structural details.

ACKNOWLEDGMENTS

Research in the Billinge group was supported by the US National Science foundation through Grant DMR-0703940. Use of the APS is supported by the US DOE, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. The 6ID-D beamline in the MUCAT sector at the APS is supported by the US DOE, Office of Science, Office of Basic Energy Sciences, through the Ames Laboratory under Contract No. W-7405-Eng-82. Beamtime on NPDF at Lujan Center at Los Alamos National Laboratory was funded under DOE Contract No. DEAC52-06NA25396.

*sb2896@columbia.edu

[1] S. J. L. Billinge and I. Levin, Science **316**, 561 (2007).
[2] C. Giacovazzo, G. Gilli, H. L. Monaco, and D. Viterbo, *Fundamentals of Crystallography* (Oxford University Press, Oxford, 1992).
[3] R. A. Young, *The Rietveld Method*, *International Union of Crystallography Monographs on Crystallography* Vol. 5 (Oxford University Press, Oxford, 1993).
[4] S. J. L. Billinge, R. G. DiFrancesco, G. H. Kwei, J. J. Neumeier, and J. D. Thompson, Phys. Rev. Lett. **77**, 715 (1996).
[5] A. Mesaros, K. Fujita, H. Eisaki, S. Uchida, J. C. Davis, S. Sachdev, J. Zaanen, M. J. Lawler, and E.-A. Kim, Science **333**, 426 (2011).
[6] I.-K. Jeong, T. W. Darling, J. K. Lee, T. Proffen, R. H. Heffner, J. S. Park, K. S. Hong, W. Dmowski, and T. Egami, Phys. Rev. Lett. **94**, 147602 (2005).
[7] E. S. Božin, C. D. Malliakas, P. Souvatzis, T. Proffen, N. A. Spaldin, M. G. Kanatzidis, and S. J. L. Billinge, Science **330**, 1660 (2010).
[8] S. J. L. Billinge, J. Solid State Chem. **181**, 1698 (2008).
[9] S. J. L. Billinge and M. G. Kanatzidis, Chem. Commun. **2004**, 749 (2004).
[10] C. A. Young and A. L. Goodwin, J. Mater. Chem. **21**, 6464 (2011).
[11] A. S. Masadeh, E. S. Božin, C. L. Farrow, G. Paglia, P. Juhás, S. J. L. Billinge, A. Karkamkar, and M. G. Kanatzidis, Phys. Rev. B **76**, 115413 (2007).
[12] S. J. L. Billinge, T. Dykhne, P. Juhás, E. Božin, R. Taylor, A. J. Florence, and K. Shankland, CrystEngComm **12**, 1366 (2010).
[13] K. W. Chapman, P. J. Chupas, E. R. Maxey, and J. W. Richardson, Chem. Commun. **2006**, 4013 (2006).
[14] M. Shatnawi, G. Paglia, J. L. Dye, K. D. Cram, M. Lefenfeld, and S. J. L. Billinge, J. Am. Chem. Soc. **129**, 1386 (2007).
[15] T. Proffen, T. Egami, S. J. L. Billinge, A. K. Cheetham, D. Louca, and J. B. Parise, Appl. Phys. A **74**, s163 (2002).
[16] P. J. Chupas, K. W. Chapman, and P. L. Lee, J. Appl. Crystallogr. **40**, 463 (2007).
[17] C. L. Farrow, P. Juhás, J. Liu, D. Bryndin, E. S. Božin, J. Bloch, T. Proffen, and S. J. L. Billinge, J. Phys. Condens. Matter **19**, 335219 (2007).
[18] X. Qiu, J. W. Thompson, and S. J. L. Billinge, J. Appl. Crystallogr. **37**, 678 (2004).
[19] M. G. Tucker, M. T. Dove, and D. A. Keen, J. Appl. Crystallogr. **34**, 630 (2001).
[20] P. F. Peterson, M. Gutmann, T. Proffen, and S. J. L. Billinge, J. Appl. Crystallogr. **33**, 1192 (2000).
[21] A. K. Soper, Chem. Phys. **202**, 295 (1996).
[22] T. Proffen and S. J. L. Billinge, J. Appl. Crystallogr. **32**, 572 (1999).
[23] B. Gilbert, F. Huang, H. Zhang, G. A. Waychunas, and J. F. Banfield, Science **305**, 651 (2004).
[24] N. J. Tosca, S. M. McLennan, M. D. Dyar, E. C. Sklute, and F. M. Michel, J. Geophys. Res. **113**, E05005 (2008).
[25] C. A. Simpson, C. L. Farrow, P. Tian, S. J. L. Billinge, B. J. Huffman, K. M. Harkness, and D. E. Cliffel, Inorg. Chem. **49**, 10858 (2010).
[26] B. H. Toby and S. J. L. Billinge, Acta Crystallogr. Sect. A **60**, 315 (2004).
[27] T. Egami and S. J. L. Billinge, *Underneath the Bragg Peaks: Structural Analysis of Complex Materials* (Pergamon, Elsevier, Oxford, England, 2003).
[28] D. Schwarzenbach *et al.*, Acta Crystallogr. Sect. A **45**, 63 (1989).
[29] E. A. Stern, Phys. Rev. B **48**, 9825 (1993).
[30] C. E. Shannon, Proc. IRE **37**, 10 (1949).
[31] C. L. Farrow and S. J. L. Billinge, Acta Crystallogr. Sect. A **65**, 232 (2009).
[32] K. Levenberg, Q. Appl. Math. **2**, 164 (1944).
[33] D. Marquardt, SIAM J. Appl. Math **11**, 431 (1963).
[34] E. Whittaker, P. Roy. Soc. Edinb. A **35**, 181 (1915).
[35] L. B. McCusker, R. B. V. Dreele, D. E. Cox, D. Louër, and P. Scardi, J. Appl. Crystallogr. **32**, 36 (1999).
[36] P. J. Chupas, X. Qiu, J. C. Hanson, P. L. Lee, C. P. Grey, and S. J. L. Billinge, J. Appl. Crystallogr. **36**, 1342 (2003).
[37] A. P. Hammersley (1998), ESRF Internal Report ESRF98HA01T.
[38] X. Qiu, Th. Proffen, J. F. Mitchell, and S. J. L. Billinge, Phys. Rev. Lett. **94**, 177203 (2005).
[39] E. S. Božin, X. Qiu, M. Schmidt, G. Paglia, J. F. Mitchell, P. G. Radaelli, T. Proffen, and S. J. L. Billinge, Physica B **385-386**, 110 (2006).