



Structure Determination of Disordered Materials from Diffraction Data

Matthew J. Cliffe,¹ Martin T. Dove,¹ D. A. Drabold,² and Andrew L. Goodwin^{1,3,*}

¹*Department of Earth Sciences, Cambridge University, Downing Street, Cambridge CB2 3EQ, United Kingdom*

²*Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701-2979, USA*

³*Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, United Kingdom*

(Received 8 December 2009; revised manuscript received 13 January 2010; published 22 March 2010)

We show that the information gained in spectroscopic experiments regarding the number and distribution of atomic environments can be used as a valuable constraint in the refinement of the atomic-scale structures of nanostructured or amorphous materials from pair distribution function (PDF) data. We illustrate the effectiveness of this approach for three paradigmatic disordered systems: molecular C₆₀, *a*-Si, and *a*-SiO₂. Much improved atomistic models are attained in each case without any *a priori* assumptions regarding coordination number or local geometry. We propose that this approach may form the basis for a generalized methodology for structure “solution” from PDF data applicable to network, nanostructured and molecular systems alike.

DOI: 10.1103/PhysRevLett.104.125501

PACS numbers: 61.43.-j, 02.70.Rr, 61.46.-w, 81.07.Nb

Many materials of fundamental importance possess structures that do not exhibit long-range periodicity: examples include metallic and covalent glasses [1], amorphous biominerals [2], the so-called “phase-change” chalcogenides of DVD-RAM technology [3], and amorphous semiconductors such as *a*-Si and *a*-Ge [4]. The absence of Bragg reflections in the diffraction patterns of these materials precludes the use of traditional crystallographic techniques as a means of determining their atomic-scale structures. Yet it is clear that these materials do possess well-defined local structure on the nanometer scale [5]; moreover it is often this local structure that is implicated in the particular physical properties of interest [6]. For this reason, the development of systematic information-based methodologies for the determination of local structure in disordered materials remains one of the key challenges in modern structural science [7].

Historically, local structure has been studied experimentally using two principal approaches: (i) the diffraction techniques of neutron and x-ray total scattering, from which the distribution of interatomic separations can be measured via the pair distribution function (PDF) [8], and (ii) resonance and spectroscopic methods (NMR, EXAFS, IR, Raman) that yield information concerning the number and population of distinct atomic environments, together with (in favorable cases) metal-coordination or molecular geometries [9,10]. These techniques afford a rich body of information, and over the past 5–10 years a number of sophisticated methods of analysis have emerged that aim to derive structural models via fitting to these experimental data. The reverse Monte Carlo (RMC) [11] and empirical potential structure refinement (EPSR) [12] methods have been used widely in the glass and amorphous materials community, while the PDFFIT [13] and “Liga” [14] meth-

ods have been applied more recently to nanostructured solids such as C₆₀ [7] and ferrihydrite [15]—systems that present similar crystallographic challenges.

There is, however, a fundamental problem: markedly different structural models can be equally consistent with the same PDF data [16]. Moreover, the task of fitting simultaneously to PDF and spectroscopic data is almost always either too computationally demanding or in fact not quantitatively possible. Taken together, these factors have meant that it is often difficult to determine the atomic-level structure of these materials, and that there is no “routine” information-based approach analogous to those for crystalline materials.

In this Letter, we show that this problem can largely be solved by using information gained via spectroscopy—the number and population of distinct atomic environments—to guide refinement of experimental PDF data. Structural refinement based on reproducing the experimental PDF alone is, in general, not sufficiently well constrained to produce models that reflect the “true” local structure in a material; however, if refinement is forced also to reflect the correct number and distribution of atom environments then convergence on the correct local structure usually follows. This approach is easily implemented and generic. Moreover, we show that successful refinement can be initiated using entirely random atomistic models and, in being driven wholly by experimental data, one avoids any other *a priori* assumptions concerning, e.g., coordination numbers or geometries. While our focus lies on proof of principle at this stage, our results show that routine information-based structure determination of disordered materials is now a viable prospect.

Our Letter is arranged as follows. We begin by describing the particular implementation of our methodology

through a “variance”-based term in the cost function used to drive PDF refinement. Three principal case studies follow: nanoparticulate C_{60} (single cluster; one atom environment), amorphous silicon (continuous network solid; one atom environment), and amorphous silica (continuous network solid; two atom environments). In all three instances we show that a conventional RMC approach fails to obtain the correct structure solution—often spectacularly—but that inclusion of the variance term is sufficient to recover almost-perfect models of material structure in each case. We conclude by discussing a number of different possible implementations of our underlying methodology.

In outlining our methodology, it is useful to consider first the simplest type of disordered material: namely, a phase that contains a single atom type and for which spectroscopy indicates a single atom environment. The existence of a single atomic environment demands that structural correlation functions calculated for different individual atoms within the material should take similar forms. In order to recast this statement with specific reference to the PDF, we first define atomic PDFs $p_j(r)$ for an atomistic model such that the “bulk” (measurable) PDF $G(r)$ corresponds to the average $\langle p(r) \rangle$ taken over all atoms j . Then the existence of a single atom environment dictates a similarity $p_j(r) \sim p_{j'}(r) \sim G(r)$ for all atoms j, j' . Whereas a standard PDF-based structure refinement would involve minimizing a function of the form

$$\chi^2 = \sum_r [\langle p(r) \rangle - G_{\text{expt}}(r)]^2, \quad (1)$$

what we would propose is the alternative cost function

$$\chi^2 = \frac{1}{N} \sum_j \sum_r [p_j(r) - G_{\text{expt}}(r)]^2. \quad (2)$$

Note that in this reformulation one obtains $\chi^2 = 0$ if and only if the model PDF matches $G_{\text{expt}}(r)$ and each individual $p_j(r)$ has the same form. It is straightforward to show that the new penalty function χ^2 of (2) is in fact equal to that of (1) plus a variance term $\chi_{\text{var}}^2 = \langle p(r)^2 \rangle - \langle p(r) \rangle^2$. What the spectroscopic result suggests is to add to a conventional PDF refinement a term that penalizes variance in local coordination environments; for this reason we are terming our approach an invariant environment refinement technique (INVERT).

In practice, the individual $p_j(r)$ for a static atomistic model consist of a series of delta functions, and in order to obtain a well-behaved variance term, it is necessary to adopt a modified formulation such as

$$\chi_{\text{var}}^2 = \frac{1}{N} \sum_j \sum_i \frac{[d_j(i) - \langle d(i) \rangle]^2}{\langle d(i) \rangle^2}, \quad (3)$$

where $d_j(i)$ measures the distance from atom j to its i th neighbor, and $\langle d(i) \rangle$ is the average such distance over all

atoms j . The term in the denominator of Eq. (3) appears in order to account for the fact that the number of neighbors at a distance d scales with d^2 [17]. The extension to multiple atom types and/or atom environments is straightforward. A separate variance term is included for each different pair of atom types (A and B , say); the form of each individual term is the same as in Eq. (3) except that the $d_j(i)$ will refer to i th neighbor of type B around the j th atom of type A , and so on.

At this point we emphasize that no assumption has been made regarding the actual distribution of neighbors around each atom—only that this distribution should be similar for equivalent atoms. Moreover, we are able to constrain the partial PDF functions for multicomponent systems despite the experimental PDF data representing a sum over these separate contributions.

We have chosen C_{60} as a simple first case study, not least because the task of determining its well-known icosahedral structure from the experimental PDF [Fig. 1(a)] has recently been highlighted as a benchmark challenge in nanostructure determination [14]. As straightforward as the task might seem, conventional RMC refinement from a random starting configuration [Fig. 1(b)] fails entirely, giving a set of small clusters that contains only a few of the real set of interatomic separations [Fig. 1(c)]. The same result is obtained even if idealized PDF data are used.

The INVERT modification exploits the experimental NMR result that C_{60} contains a single C environment [18]. Clearly the RMC configuration in Fig. 1(c) violates

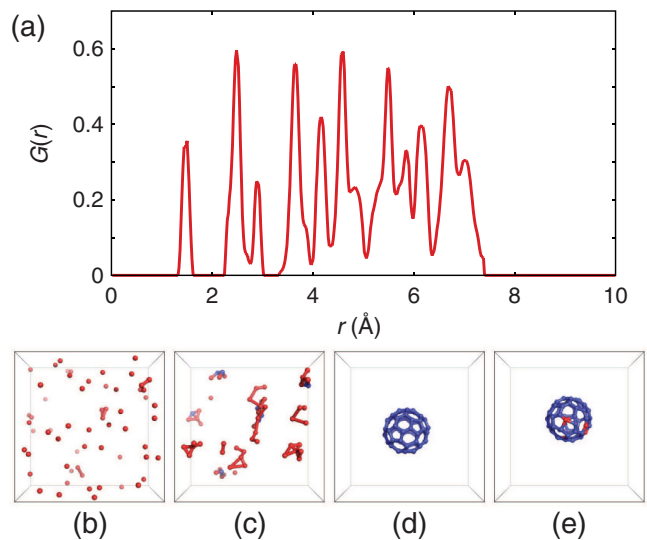


FIG. 1 (color online). RMC refinement of the experimental PDF of C_{60} : (a) the neutron PDF itself, corrected to remove intermolecular correlations as described in Ref. [14]; (b) a random starting configuration of 60 carbon atoms; (c) a typical configuration produced by conventional RMC refinement of either idealized or experimental PDF data; and those produced by INVERT + RMC using (d) idealized, and (e) experimental PDF data.

this property, and so would now give rise to a large χ_{var}^2 term that will help drive refinement forward. Indeed, INVERT + RMC refinement from the same random starting configuration gives the correct solution for idealized data [Fig. 1(d)] and a near-perfect solution for the experimental neutron PDF data of Ref. [14] [Fig. 1(e)]. We note that such a result has only ever been achieved previously using the highly sophisticated cluster optimization methods of the Liga algorithm or using genetic algorithms based on the principle of mating or crossover (the latter only giving correct solutions in 56% of attempts) [7,19,20]. Here, INVERT + RMC consistently obtains a topologically identical solution from random starting coordinates in approximately 2000–4000 accepted moves.

We find the extension to a cluster with two atom environments—namely, S_{12} [21]—enjoys similar success [22]. Videos that illustrate the refinement process for C_{60} and S_{12} are provided as supporting information [23].

The paradigmatic “stumbling block” for RMC, however, has always been amorphous Si, whose structure is believed to consist of a continuous random network (CRN) of tetrahedral Si centers [24]. Rather than generating a network of fourfold-coordinated Si atoms, RMC refinements of *a*-Si PDF data yield configurations with unphysically broad distributions of Si coordination numbers [16]. This is allowed because, crudely speaking, a pair of atoms of coordination numbers three and five will contribute to the average PDF indistinguishably from two fourfold-coordinated atoms, and yet the former state is statistically more likely during a sequence of random moves. Various work-arounds have been proposed and implemented (e.g., constraining coordination numbers to equal four), and in the most favorable of cases these yield CRNs comparable to those obtained from bond switching (Wooten-Weaire-Winer, “WWW” [25]) methods, molecular dynamics and *ab initio* calculations [24].

However, there is a sense in which one recovers from these approaches only the very information already used to generate the constraints: if the coordination number is constrained to equal four during refinement, then fourfold coordination cannot be considered an independent result. Consequently, our motivation for considering *a*-Si as a second case study was primarily to determine whether, by using the evidence for a single Si environment from NMR studies [26], INVERT + RMC refinement could yield reasonable structural models without recourse to explicit coordination number constraints.

First, a conventional RMC refinement was performed using $G(r)$ “data” generated from the trusted WWW model of Ref. [25]. The starting configuration was a random collection of 512 Si atoms in a cubic box of side 21.7 Å. Refinement gave a highly disordered configuration that displayed all the hallmarks of previously described problematic RMC studies [16]: only 27% of Si atoms are fourfold coordinated, there are substantial density varia-

tions, and large numbers of unphysical Si_3 “triangles” [left-hand panel of Fig. 2(a)]. In contrast, a parallel INVERT + RMC refinement achieved an almost perfect coordination distribution (95% fourfold). The improvement extended even to the higher-order correlations (discussed in more detail below): in particular, the number of Si_3 triangles is halved, and the density distribution is much more even. Inspection of the configuration itself [center panel of Fig. 2(a)] now reveals obvious similarities to the trusted WWW model [right-hand panel of Fig. 2(a)]. The PDF itself is relatively insensitive to this fundamental improvement in local structure modeling [Fig. 2(b)], while the variance term of Eq. (3) clearly acts as a much better figure of merit [Fig. 2(c)].

Similar results are obtained for amorphous SiO_2 , which is a conceptual extension in that it contains two distinct atom environments: that of the Si atoms and that of the O atoms. Experimental neutron PDF data were taken from Ref. [27], and starting configurations generated from a random distribution of 64 Si atoms and 128 O atoms in a periodic cubic box of side 14.37 Å. RMC refinement both with and without the INVERT modification gave excellent fits to the PDF, but the INVERT + RMC model had a much higher percentage of fourfold Si coordination

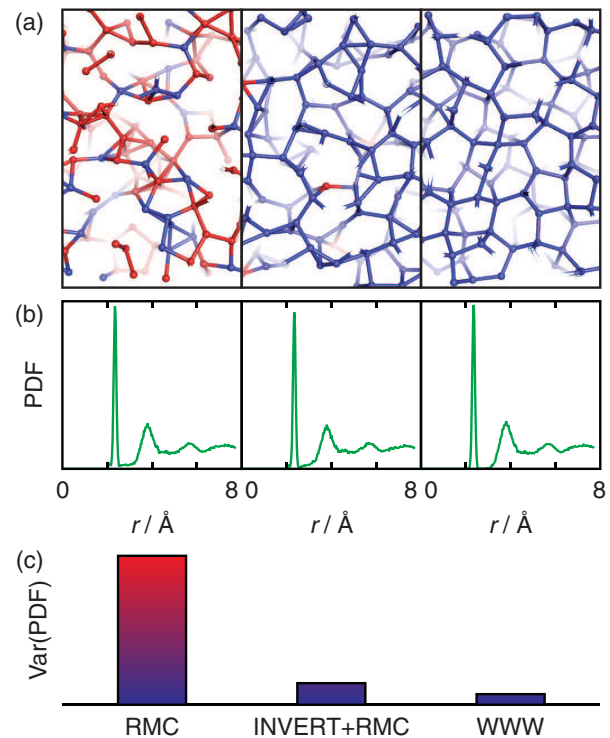


FIG. 2 (color online). Comparison of *a*-Si configurations obtained using (left) native RMC and (center) “INVERT + RMC” refinement with (right) the trusted “WWW” model of Ref. [25]. (a) Slices of the configurations themselves. (b) The PDFs calculated from each configuration, which are essentially identical. (c) Corresponding PDF variances calculated using Eq. (3).

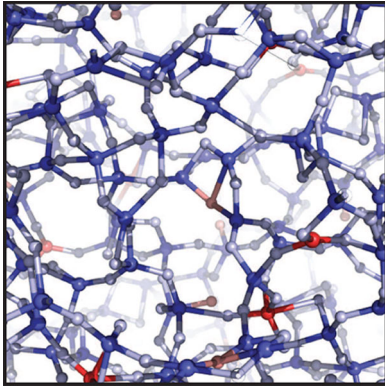


FIG. 3 (color online). A slice of the information-driven CRN model of a -SiO₂ obtained using INVERT + RMC refinement as described in the text. Si and O atoms shown in dark and light shades, respectively.

(97% vs 59% for the RMC-only configuration). Indeed, we believe the INVERT + RMC configuration to be the first information-based CRN model of a -SiO₂ [Fig. 3].

The INVERT methodology is by no means applicable only to RMC refinement. Our focus on the RMC method in this Letter arises from a desire to demonstrate the effectiveness of the INVERT approach using a refinement method that is known to favor *disorder*. The incorporation of variance-based cost functions in any refinement approach is straightforward, and such a modification to more sophisticated PDF fitting approaches than the RMC method, e.g., as suggested in Ref. [14], might reasonably be expected to produce even more realistic configurations.

Speaking more generally, we would note that the concept of local invariance encompasses more than minimizing the PDF variance alone. One can imagine, for example, that the minimizing the variance in higher-order correlation functions, such as angle distributions, coordination geometry, and CRN ring statistics might also improve refinement further. Importantly, these constraints can be implemented despite the functions not being measurable experimentally. In practice, however, we have found that the calculation of higher-order correlation functions is too computationally demanding for speedy refinement at this stage; the extension to constraining geometric invariance using spherical harmonics and/or the triplet distribution function is an approach we hope to pursue further in the near future.

We gratefully acknowledge financial support from Trinity College, Cambridge to A. L. G., from the EPSRC (U.K.) to A. L. G. and M. J. C., and from the U.S. NSF to D. A. D. under Grant No. DMR 09-03225. We thank D. A. Keen (Rutherford Appleton Laboratory) for useful discussions, and acknowledge the University of Cambridge's CamGrid infrastructure for computational resources.

*andrew.goodwin@chem.ox.ac.uk

- [1] C. J. Byrne and M. Eldrup, *Science* **321**, 502 (2008).
- [2] S. Weiner, I. Sagi, and L. Addadi, *Science* **309**, 1027 (2005).
- [3] Z. M. Sun, J. Zhou, and R. Ahuja, *Phys. Rev. Lett.* **96**, 055507 (2006).
- [4] G. S. Armatas and M. G. Kanatzidis, *Nature (London)* **441**, 1122 (2006).
- [5] S. J. L. Billinge and I. Levin, *Science* **316**, 561 (2007).
- [6] S. J. L. Billinge and M. G. Kanatzidis, *Chem. Commun. (Cambridge)* **7** (2004) 749.
- [7] P. Juhás *et al.*, *Nature (London)* **440**, 655 (2006).
- [8] T. Egami and S. J. L. Billinge, *Underneath the Bragg Peaks: Structural Analysis of Complex Materials* (Pergamon, Oxford, 2003).
- [9] M. H. Brodsky and M. Cardona, *J. Non-Cryst. Solids* **31**, 81 (1978).
- [10] W. Mullerwarmuth and H. Eckert, *Phys. Rep.* **88**, 91 (1982).
- [11] R. L. McGreevy, *J. Phys. Condens. Matter* **13**, R877 (2001).
- [12] A. K. Soper, *Phys. Rev. B* **72**, 104204 (2005).
- [13] T. Proffen and S. J. L. Billinge, *J. Appl. Crystallogr.* **32**, 572 (1999).
- [14] P. Juhás *et al.*, *Acta Crystallogr. Sect. A* **64**, 631 (2008).
- [15] F. M. Michel *et al.*, *Science* **316**, 1726 (2007).
- [16] O. Gereben and L. Pusztai, *Phys. Rev. B* **50**, 14136 (1994).
- [17] We note that other modifications will be possible—e.g., convoluting the static $p_j(r)$ with a Gaussian blurring function to simulate the effect of thermal motion—and indeed in nonstatic models (such as molecular dynamics) may not be necessary at all. We focus here on the modification given in Eq. (3) because it is easily implemented and remains a good measure of the variance.
- [18] R. D. Johnson, G. Meijer, and D. S. Bethune, *J. Am. Chem. Soc.* **112**, 8983 (1990).
- [19] B. Hartke, *J. Comput. Chem.* **20**, 1752 (1999).
- [20] D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.* **75**, 288 (1995).
- [21] J. Steidel, R. Steudel, and A. Kutoglu, *Z. Anorg. Allg. Chem.* **476**, 171 (1981).
- [22] The molecule S₁₂ contains six atoms in each of two distinct environments [21]. PDF refinements started from a random arrangement of 6 “S_A” and 6 “S_B” atoms, and the variance contribution taken to equal the sum of the four partial variance terms (S_A–S_A, S_A–S_B, S_B–S_A, S_B–S_B). The correct molecular structure was consistently attained within 1000–2000 accepted moves; RMC refinements without the INVERT modification gave unphysical clusters not dissimilar in form to those obtained for C₆₀.
- [23] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.104.125501> for videos showing the refinement of the structures C₆₀ and S₁₂.
- [24] D. A. Drabold, *Eur. Phys. J. B* **68**, 1 (2009).
- [25] F. Wooten, K. Winer, and D. Weaire, *Phys. Rev. Lett.* **54**, 1392 (1985).
- [26] W.-L. Shao *et al.*, *Phys. Rev. B* **41**, 9491 (1990).
- [27] M. G. Tucker *et al.*, *J. Phys. Condens. Matter* **17**, S67 (2005).