



## PAPER

## OPEN ACCESS

## RECEIVED

24 January 2020

## ACCEPTED FOR PUBLICATION

12 February 2020

## PUBLISHED

6 March 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# Generating large starting configurations for molecular Reverse Monte Carlo modelling of an unique non-linear optical amorphous solid

B D Klee<sup>1,4</sup> , B Paulus<sup>1</sup>, S Hosokawa<sup>1,2</sup> , M T Wharmby<sup>3</sup> , E Dornsiepen<sup>1</sup> , S Dehnen<sup>1</sup> and W-C Pilgrim<sup>1</sup>

<sup>1</sup> Department of Chemistry, Philipps-University, Hans-Meerwein Str. 4, 35032 Marburg, Germany

<sup>2</sup> Department of Physics, Kumamoto University, Kumamoto 860-8555, Japan

<sup>3</sup> PETRA III, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany

<sup>4</sup> Author to whom any correspondence should be addressed.

E-mail: [benjamin.klee@staff.uni-marburg.de](mailto:benjamin.klee@staff.uni-marburg.de) and [pilgrim@staff.uni-marburg.de](mailto:pilgrim@staff.uni-marburg.de)

**Keywords:** Reverse Monte Carlo, amorphous structure, X-Ray scattering, white-light generation, molecular solids

## Abstract

We report on the recent advances regarding the source code optimization of Reverse Monte Carlo modelling used in scattering data analysis of an amorphous molecular solid which has recently attracted attention as a new brilliant white light emitter if irradiated by a simple infrared laser diode. The algorithm used for generating random molecular starting configurations without overlapping molecules in a box with periodic boundary conditions has been accelerated by a factor of roughly 400 in a 54k atom case. The resulting bigger independent starting configurations are used to gain further insight into previously presented x-ray scattering data. New improved scattering data have been obtained, revealing new structural features in the lower  $Q$  range.

## Introduction

Recently, a new class of materials has been synthesized which exhibits unique non-linear optical properties. Some of these substances convert infrared light from a commercially available low-cost low-power laser diode into brilliant white light, thereby maintaining the favourable coherent emission characteristics of the laser [1–5]. This so-called supercontinuum generation was so far only possible if expensive high-power pulsed laser sources were used, and was therefore mainly restricted to academic applications [6–8]. These new materials consist of semiconductor-based cluster molecules with general molecular formula  $[(RT)_4S_6]$  with  $T = \text{Ge or Sn}$  and  $R$  being a covalently bound organic group like e.g., phenyl or steryl. Interestingly, none of these substances could yet be prepared in a decent crystalline state. Instead, they are only available as disordered amorphous solids or show at the most marginal crystallization tendency. Pure crystallinity was found in some related materials like e.g.  $[(\text{PhSi})_4S_6]$  where Sn was replaced by Si and where phenyl substituents were again chosen as suitable organic substituents. Instead of supercontinuum generation however, these crystalline species were found to be highly efficient second harmonics (SHG) generators. Structural disorder seems hence to be one prerequisite for the supercontinuum generation in these materials [2]. A vivid physical explanation for this effect does not yet exist. In the current understanding,  $\pi$ -electrons of the organic ligands are driven into the cluster's ground state potential minimum by interaction with IR-radiation, and the released energy is emitted as white light [1]. The apparent need of disorder suggests that specific structural arrangements between molecular entities may exist which cannot be engendered in the crystalline state. A detailed perception of the cluster arrangements in the amorphous phases may therefore provide key information to fully understand this new fascinating effect.

In order to thoroughly elucidate such structural correlations, we have recently started a comprehensive scattering study using synchrotron radiation. However, a reliable interpretation of the resulting total multi-component pair correlation functions is virtually impossible and therefore the use of Reverse Monte Carlo (RMC) techniques is in general appropriate. In classical RMC, computer-generated sample atoms inside a virtual simulation box are iteratively moved, each time calculating the resulting scattering law  $S(Q)_{\text{calc}}$  based on



the atom positions, which is compared with the experimental  $S(Q)$ . This process is repeated until satisfying agreement between calculated and experimental  $S(Q)$  is reached and the calculation is stopped. The virtual simulation box may then contain a suitable copy of the real atomic arrangement in the sample and the individual partial pair correlation functions can be computed. From these functions and from the virtual atomic arrangement in the simulation box, one may then infer the desired specific microscopic structural correlations [9]. However, this procedure is ambiguous and may lead to several possible atomic arrangements which conform to the same experimental  $S(Q)$  [10]. Hence, additional constraints are needed to improve the reliability of RMC and to enforce unique and realistic simulation results. Minimum distances between the atoms based on the individual atomic sizes act as first indispensable constraints. Additionally, different scattering laws of the same substance are often employed. Such functions may be available from neutron scattering using isotopic substitution techniques [11–13] or from anomalous x-ray scattering [12, 14, 15]. Another strong constraint exists if known significant structural correlations exist between certain atoms in the sample as is the case in disordered materials made up of molecular units. It appears appropriate to move whole molecular units instead of single atoms during the RMC-iterations to intrinsically preserve the molecular structure during the simulation. There have been RMC studies on molecular liquids before in which groups of atoms constituting single molecular units were defined and these entities were moved and rotated as a whole during the iteration steps [16, 17]. Such studies were however mostly limited to small molecules which could be assumed as being rigid.

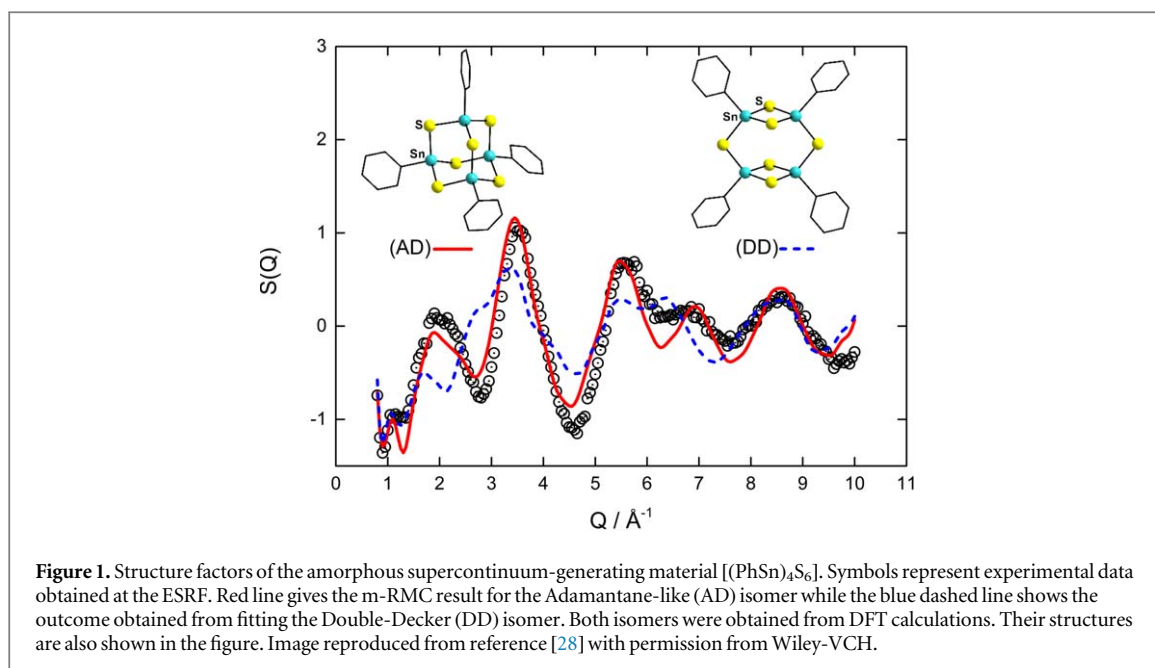
The implementation of Molecular RMC-Simulations (m-RMCS) is difficult and comprises several disadvantages: (1) the low mobility of the molecules due to the likelihood of collisions with other molecules during a single iteration leads to slow convergence of the fitting procedure. (2) prior knowledge of molecular structure is needed and (3) the uncertainty regarding the atom positions due to thermal vibrations is difficult to include. The latter is usually taken care of by introducing Debye-Waller factors without taking into account the additional space needed by the atoms due to the vibrations, virtually leading to an unrealistically large free volume [18]. Furthermore the isotropic broadening of the atom positions introduced this way may not be sufficient to describe the deformation of molecules that usually leads to a more ellipsoid-like distribution of atom positions in respect to the rest of the molecule. These problems led to the conclusion that in general atom based moves are more efficient for rigid molecules [19]. Instead of rigidly defined molecules in m-RMCS nowadays simulations based on molecular dynamics (MD) are employed to explore microscopic correlations in amorphous or liquid molecular systems [20–22]. The MD approach uses force-field parameters to model the interatomic potentials and defines molecules in a non-rigid way. Available sets of force-field parameters are typically limited to chemically similar classes of substances and distinguish atoms based on their chemical environment [23, 24]. For some substances like water and biologically relevant molecules several sets of well tested force-field parameters exist [25, 26] but no force-field set has yet been developed to describe the bonding situation in the systems considered here.

In this paper, we present another attempt to analyse x-ray scattering data obtained from one of these new fascinating amorphous molecular systems using m-RMCS. We report about recent improvements of the simulation code which led to a considerable enhancement of simulation speed. For this we have modified a conventional RMC-Code based on the RMC\_POT++ program code of Gereben *et al* [27] which already offers the option to define groups of atoms within a custom function which can then be moved as molecular units. One of the main challenges along this strategy is however to optimize involved algorithms such as the generation of suitable starting configurations. Such configurations must contain the desired number of intact molecules, respect the user-defined minimum distances, feature periodic boundary conditions and have an atom density representing the macroscopic density of the sample. Starting configurations can be obtained from known crystal structures or, in case of a standard RMC approach, by moving randomly placed atoms away from each other until all minimum distances are reached. In case of larger molecules with a distinct shape finding a sufficiently large random starting configuration becomes non-trivial and no standard routine code did so far exist to allow the generation of such starting configurations on reasonable time scales.

## Experimental

A sample of the amorphous white light emitter  $[(\text{PhSn})_4\text{S}_6]$  was prepared according to literature procedures [28] and first scattering data were obtained at the European Synchrotron Radiation Facility (ESRF) at beamline BM02 [29]. Primary energy at that beamline was set to 28.9 keV and measurements were performed in transmission geometry. The powdered sample was confined in a custom-made brass cell between two 50  $\mu\text{m}$  Kapton windows. Sample thickness was set to 1 mm due to the low x-ray cross section of the sample. Usual data reduction procedures were performed on the raw data to account for background and absorption and data were normalized to density and form factors. The resulting  $S(Q)$  is given in figure 1 by the symbols. The statistical





quality of these data is lower than in usual scattering experiments at synchrotron sources. This is due to the high energy resolution of the setup which is optimized for anomalous x-ray scattering experiments paired with a very short data acquisition time that was limited due to the onset of sample decomposition in this energy range.

A sample of the same material was also used in another diffraction experiment carried out quite recently on beamline P02.1 at the Petra III-Ring of the Deutsche Elektron Synchrotron, DESY. There, the sample was confined in a 1 mm x-ray capillary for transmission measurements. A primary energy of 59.87 keV was used and scattering data were collected using a flat pixel detector with pixel size  $200 \times 200 \mu\text{m}^2$  and a sample detector distance of 240.18 mm.

Analysis of the ESRF-data were performed using our modified m-RMCS code based on the RMC\_POT++ program of Gereben *et al* [27]. Knowledge on molecular structure was obtained from density functional theory (DFT) calculations of the single molecule [2]. An Adamantane-like (AD) structure as well as a Double-Decker (DD) structure were obtained from this study however, with an energetical preference for the AD-type. Both isomers are displayed in figure 1. Custom made structural functions were written for the two different  $[(\text{PhSn})_4\text{S}_6]$  clusters which allowed translational and rotational moves of the complete units as well as individual rotation of the phenyl groups [28]. Vibrational motions of the respective atoms were simulated by convoluting the acquired pair correlation functions with a suitable gaussian. Random starting configurations without overlapping molecules and an appropriate density were created from scratch in order to run the m-RMCS. Best fits achieved for both of the isomers are shown in figure 1 as the blue dashed and the red solid line, respectively. It is obvious that the general features of  $S(Q)$  are much better reproduced by the AD isomer which agrees with the energetical prediction from the DFT-calculations.

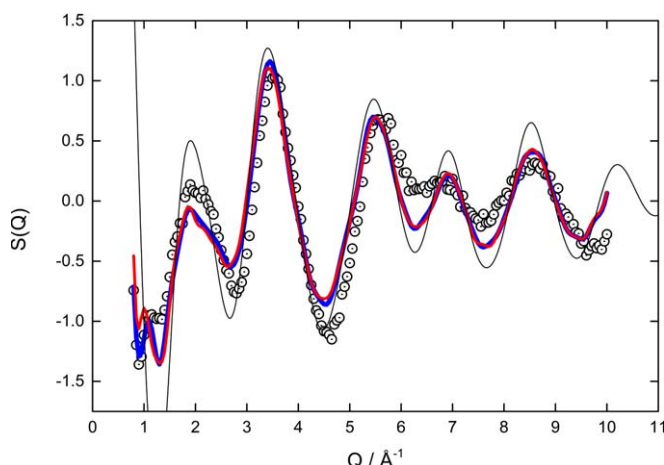
The original so called moveout option of the RMC\_POT++ package which is used to generate starting configurations of atomic systems is not designed to disentangle randomly placed and therefore overlapping molecules. Therefore this routine was altered to enable the creation of starting configurations containing complex molecules [28]. The central quantity in this new routine is the so called moveout sum  $\mathbb{S}$ , it is given as

$$\mathbb{S} = \sum_i^N \sum_{j \geq i}^N \Delta d_{i,j},$$

and contains contributions  $\Delta d_{i,j}$  of atom pairs which are closer than the respective user defined cut-off constraints. For m-RMCS this is also a measure for quantifying the amount of overlap between different molecules. Random moves, which minimize  $\mathbb{S}$  will disentangle the molecules and result in a configuration where all user-defined cutoff-constraints are satisfied.  $\mathbb{S}$  is recalculated every time a molecule is moved which leads to bad scaling with increasing number of atoms due to the double sum over all atoms  $N$ . Therefore, only random configurations containing 8 molecules (432 atoms) could at that stage be generated in reasonable computing time. [28] These sets of 8 molecules were then copied and reproduced to generate the full starting configuration containing 216 molecules.

To tackle the scaling problem the calculation of  $\mathbb{S}$  makes now use of simulation box gridding, where the simulation box is split in smaller sub-boxes and it is kept track which atom is located in which sub-box. The





**Figure 2.** Comparison of the m-RMCS results using 8 (blue line) and 216 molecules (red line) to generate the starting configuration. The thin black line is a calculation of the structure factor given by a single AD-molecule in an otherwise empty simulation box. Symbols represent the experimental ESRF-data.

computation of the moveout sum now reads

$$\mathbb{S} = \sum_i^N \sum_j^{N_{grid}} \Delta d_{i,j}.$$

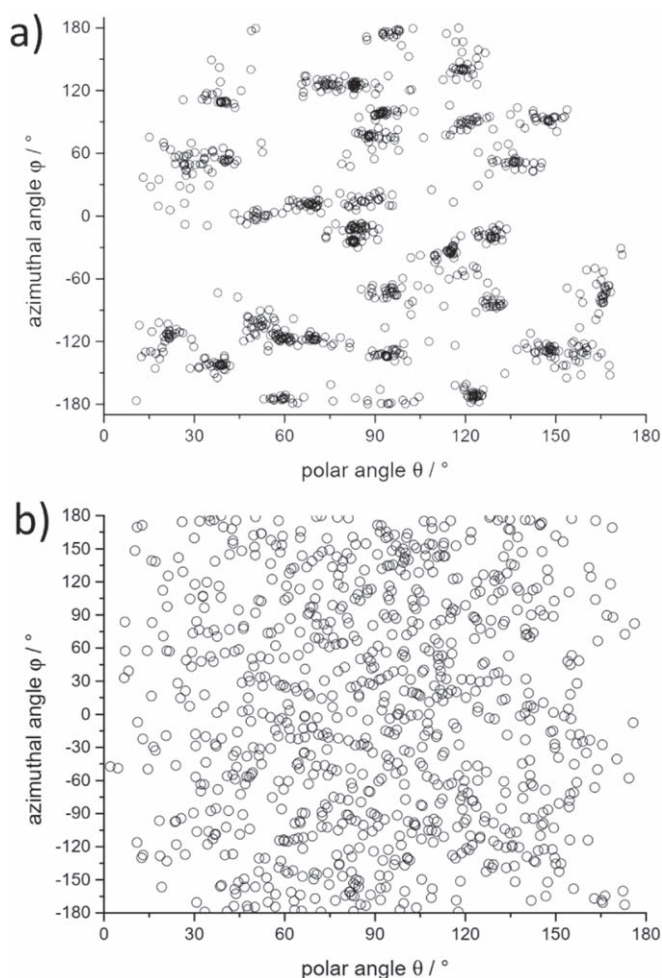
The option to perform simulation box gridding was already implemented into the original code of the RMC\_POT++ program [30]. The advantage of this definition is that the second sum does no longer run over all  $N$  atoms but only over  $N_{grid}$  atoms which are located in the neighbouring grid cells where  $N_{grid}$  is not increasing with  $N$ . If the size of the grid cells is chosen to be slightly larger than the largest user-defined cut-off constraint, then only directly neighbouring grid cells are scanned by the program and the calculation speed is significantly boosted. Another substantial upgrade of the algorithm is the implementation of multithreading for the computation of  $\mathbb{S}$ . The new implementation lets each thread calculate one element of the sum over  $i$  and each time a thread finishes it starts with the next element. Also, in the previous version molecules were chosen and moved randomly until a proper starting configuration was found. With increasing number of molecules this led to a lot of wasted time when moving molecules that were already disentangled. To avoid this, the algorithm now keeps track which pair of molecules contributes the largest  $\Delta d_{i,j}$  to  $\mathbb{S}$  and chooses one of these for the next move with increased probability. The average time needed for a move in the later stages of disentangling  $[(\text{PhSn})_4\text{S}_6]$ -molecules can be compared with the old version of the algorithm: In a 216 molecule (11664atoms) case the speed was improved from 433 ms/move to 14 ms/move and in a 1000 molecule case (54000atoms) from 9257 ms/move to 22 ms/move. All simulations for this comparison were run on the same computer using 6 threads and similar sets of parameters. Using the new code, starting configurations containing 216 molecules have been generated from a random configuration and were used to perform new m-RMCS on the above mentioned ESRF-data. The result is shown in figure 2 as the red line where it is compared with the preceding simulation result where just 8 molecules could be used to generate the starting configuration (blue line).

## Discussion

The agreement of both results shown in figure 2 with the experimental data is obviously not perfect. Differences in the peak positions and amplitudes are visible, which is due to the fact that the molecules were regarded as rigid during the simulation. In the real structure they may be distorted due to interactions with neighbouring molecules, leading to slight deviations of the average atom positions. The scattering pattern of disordered molecular systems is dominated by intramolecular structural correlations given by the molecule structure, especially at higher  $Q$  [31, 32]. The contribution of the molecule structure was estimated by calculating  $S(Q)$  of a single rigid molecule in a large empty simulation box using the gaussians to account for vibrations as for the other m-RMCS. This contribution is also shown in figure 2 as thin black line and is apparently responsible for the general features of the experimental pattern.

The two simulation results are basically identical. Small differences are only visible in the lower  $Q$ -range below about  $2.5 \text{ \AA}^{-1}$  and may be caused by slightly different relative intermolecular structural correlations in the resulting data sets. We had in fact found intermolecular correlations concerning the relative orientation between the molecules employing the 8-molecule starting configuration [28]. Such correlations were described in the



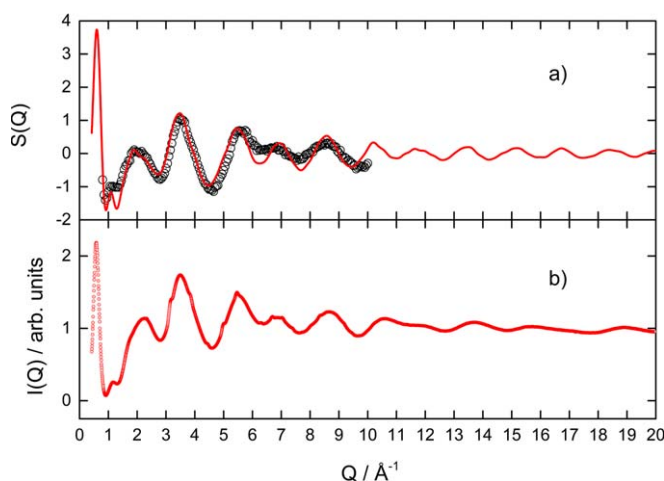


**Figure 3.** Each circle marks the azimuthal and polar angle of the vector representation of a tin-carbon bond in spherical coordinates. Panel a) shows the cluster orientations in the best fitting for the 8 molecules case while panel b) shows the cluster orientations for the simulation with 216 molecules.

following way: each tin-carbon bond in the simulation box has been extracted as a vector pointing away from the molecule centre. Hence, each molecule contributes four vectors in a tetrahedral geometry. The directions in which these vectors point can be expressed as a set of polar ( $\theta$ ) and azimuthal ( $\varphi$ ) angles in a spherical coordinate system. The result of the m-RMCS with the 8-molecule starting configuration is displayed in figure 3(a). Different vector orientations group together, indicating several preferred absolute orientations among the molecules in the simulation box. The same procedure was applied to the result obtained from the 216-molecule starting configuration and the result is shown in figure 3(b). Apparently, no characteristic absolute orientation correlation can be observed any more. Hence the effect may have been a statistical artefact related to the fact that the original starting configuration consisted of 27 boxes each containing 8 molecules with identical orientations. It has to be noted however, that this does not exclude preferential relative orientations of neighbouring molecules which are not necessarily visible when plotting the absolute orientations in an arbitrarily oriented coordinate system.

Since the ESRF-data in the figures were only available along a limited  $Q$ -range, we have re-calculated the structure factor in a wider range using the atomic positions obtained from the 216 particle m-RMCS. The result is displayed in figure 4(a). A strong pre-peak located at about  $0.59 \text{ \AA}^{-1}$  now appears as an unexpected feature. It is clearly not resulting from intramolecular structural correlations and may therefore be an important feature for the understanding of the molecular alignment in the disordered solids. It should be mentioned that we have recently succeeded in repeating the above mentioned ESRF-experiment on beamline P02.1 at the DESY facility in Hamburg [33, 34]. The scattering pattern obtained from this experiment is displayed in figure 4(b). We were not yet able to properly correct these data with respect to air scattering, absorption and Compton-contributions, hence raw data in arbitrary units are displayed. Nevertheless, figure 4 shows that these data reproduce the ESRF results very well however, extending along a considerably wider  $Q$ -range and featuring much better statistics. The important point to note here is, that the low- $Q$  peak found from the recalculated ESRF-data is present in





**Figure 4.** Panel a) shows the recalculated extended  $S(Q)$  (red line) resulting from fitting the ESRF-scattering data (black circles). New scattering data (small red circles) obtained at the P02.1 beamline at DESY featuring a larger  $Q$ -range are shown below in panel b). Image partly reproduced from reference [28] with permission from Wiley-VCH.

these experimental results too. These additional data are likely mandatory to extract reliable information regarding the relative orientation of neighbouring molecules during further investigations.

## Conclusion

Source code optimizations for rigid molecular simulations have led to a significant increase of calculation speed. This enables the generation of larger random starting configurations for molecular Reverse Monte Carlo modelling, thus eliminating statistical artefacts. New scattering data are presented and reveal a large signal in the low- $Q$  region which may be a fingerprint of intermolecular correlations in the disordered molecular solid.

## Acknowledgments

This work is funded by the Deutsche Forschungsgemeinschaft (DFG) in the scope of project FOR 2824. This publication makes use of powder diffraction and total scattering data measured as part of the rapid access program available at beamline P02.1, PETRA III at DESY (Hamburg Germany), a member of the Helmholtz Association HGF. Furthermore this publication makes use of x-ray scattering data obtained in the scope of proposal HC-3146 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The research leading to this result has been supported by the project CALIPSOplus under the Grant Agreement 730872 from the EU Framework Programme for Research and Innovation HORIZON 2020.

## ORCID iDs

B D Klee  <https://orcid.org/0000-0002-4121-9995>  
S Hosokawa  <https://orcid.org/0000-0002-7810-6062>  
M T Wharmby  <https://orcid.org/0000-0002-1764-0501>  
E Dornsiepen  <https://orcid.org/0000-0002-8711-7361>  
S Dehnen  <https://orcid.org/0000-0002-1325-9228>  
W-C Pilgrim  <https://orcid.org/0000-0003-0176-5729>

## References

- [1] Rosemann N W, Eussner J P, Beyer A, Koch S W, Volz K, Dehnen S and Chatterjee S 2016 *Science* **352** 1301–4
- [2] Rosemann N W, Eußner J P, Dornsiepen E, Chatterjee S and Dehnen S 2016 *JACS* **138** 16224–7
- [3] Rosemann N W, Locke H, Schreiner P R and Chatterjee S 2018 *Adv. Opt. Mater.* **6** 1701162
- [4] Dornsiepen E, Dobener F, Mengel N, Lenchuk O, Dues C, Sanna S, Mollenhauer D, Chatterjee S and Dehnen S 2019 *Adv. Opt. Mater.* **7** 1801793
- [5] Dornsiepen E, Dobener F, Chatterjee S and Dehnen S 2019 *Angew. Chem. Int. Ed.* **58** 17041–6
- [6] Alfano R R 2016 *The Supercontinuum Laser Source* (New York: Springer-Verlag) (<https://doi.org/10.1007/978-1-4939-3326-6>)
- [7] Dudley J M, Genty G and Coen S 2006 *Rev. Mod. Phys.* **78** 1135–84



- [8] Alfano R R and Shapiro S L 1970 *Phys. Rev. Lett.* **24** 584–7
- [9] McGreevy R L and Howe M A 1992 *Annu. Rev. Mater. Sci.* **22** 217–42
- [10] McGreevy R L 1995 *Nuclear Instruments and Methods in Physics Research section A: Accelerators Spectrometers Detectors and Associated Equipment* **354** 1–16
- [11] Salmon P S, Xin S and Henry E 1998 *Phys. Rev. B* **58** 6115–23
- [12] Fischer H E, Barnes A C and Salmon P S 2005 *Rep. Prog. Phys.* **69** 233–99
- [13] Zeidler A, Salmon P S, Whittaker D A J, Piarristeguy A, Pradel A, Fischer H E, Benmore C J and Gulbiten O 2018 *Royal Society Open Science* **5** 171401
- [14] Hosokawa S, Pilgrim W-C, Béar J-F and Boolchand P 2016 *J. Non-Cryst. Solids* **431** 31–5
- [15] Stellhorn J R, Hosokawa S, Pilgrim W-C, Blanc N, Boudet N, Tajiri H and Kohara S 2015 *Physica Status Solidi (b)* **253** 1038–45
- [16] Scheidler M, North A N and Dore J C 1993 *Mol. Simul.* **11** 345–63
- [17] Howe M 1990 *Mol. Phys.* **69** 161–74
- [18] Soper A K 2001 *Mol. Phys.* **99** 1503–16
- [19] McGreevy R L 2001 *J. Phys. Condens. Matter* **13** R877–913
- [20] Hernández E R 2008 *AIP Conference Proceedings* **95**–123
- [21] Nilsson A and Pettersson L 2011 *Chem. Phys.* **389** 1–34
- [22] Bopp P A, Buhn J B, Maier H A and Hampe M J 2008 *Chem. Eng. Commun.* **195** 1437–56
- [23] Kumari I, Sandhu P, Ahmed M and Akhter Y 2017 *Current Protein & Peptide Science* **18** 1163–79
- [24] Jorgensen W L and Tirado-Rives J 2005 *Proc. Natl Acad. Sci.* **102** 6665–70
- [25] Jorgensen W L and Tirado-Rives J 1988 *JACS* **110** 1657–66
- [26] Jorgensen W L, Maxwell D S and Tirado-Rives J 1996 *JACS* **118** 11225–36
- [27] Gereben O and Pusztai L 2012 *J. Comput. Chem.* **33** 2285–91
- [28] Klee B D, Dornsiepen E, Stellhorn J R, Paulus B, Hosokawa S, Dehnen S and Pilgrim W-C 2018 *Physica Status Solidi (b)* **255** 1800083
- [29] Ferrer J L, Simon J P, Berar J F, Caillot B, Fanchon E, Kaikati O, Arnaud S, Guidotti M, Pirocchi M and Roth M 1998 *J. Synchrotron Radiat.* **5** 1346–56
- [30] Gereben O 2015 RMC\_POT user guide for version 1.4, last checked 06.03.2019 (<http://szfki.hu/~nphys/rmc++/docs.html>)
- [31] Page D and Powles J 1971 *Mol. Phys.* **21** 901–26
- [32] Powles J 1973 *Adv. Phys.* **22** 1–56
- [33] Dippel A-C, Liermann H-P, Delitz J T, Walter P, Schulte-Schrepping H, Seeck O H and Franz H 2015 *J. Synchrotron Radiat.* **22** 675–87
- [34] Filik J et al 2017 *J. Appl. Crystallogr.* **50** 959–66