Science & Technology Facilities Council

Technical Report
RAL-TR-2017-014

Optimising molecular dynamic potentials to dynamical structure factor data

AJ Markvardsen, RL McGreevy

December 2017

Enquiries concerning this report should be addressed to:

RAL Library
STFC Rutherford Appleton Laboratory
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
Fax: +44(0)1235 446403
email: libraryral@stfc.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk

ISSN 1358-6254

Optimising molecular dynamic potentials to dynamical structure factor data

Anders J. Markvardsen and Robert L. McGreevy

ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, UK

Abstract

This paper reports an algorithm and software for producing dynamical models of materials at the atomic scale using classical molecular dynamics simulation, but simultaneously refining the potential energy parameters based on experimentally determined dynamical and structural information such as the dynamical structure factor S(Q, ω). The software is called MDMC (Molecular Dynamics - Monte Carlo) (https://github.com/MDMCproject) and the first pre-release version used in this work is https://doi.org/10.5281/zenodo.1068365.

The algorithm follows from two previous very successful algorithms Reverse Monte Carlo (RMC) (McGreevy & Pusztai, 1988)(McGreevy, 2001) and Empirical Potential Structure Refinement (EPSR) (Soper, 1996) (Soper, 2007), though these are only used to produce structural models based on structural data.

In this paper the algorithm is presented together with initial results for a model system, liquid argon, based on the data of (van Well, Verkerk, de Graaf, Suck, & Copley, 1985).

Introduction

Both RMC and EPSR have had a large impact on structural studies of disordered and amorphous materials. For recent reviews and works using these methods see for example: (Playford, Owen, Levin, & Tucker, 2014) (Bouty, 2014) (Soper & Edler, 2017). In RMC the atomic coordinates in a structural model, typically many thousands of atoms, are refined based on the difference between experimental structural data and that calculated from the model. Sequences of RMC configurations (models) have been used to explore dynamical aspects of structures (see e.g. McGreevy and Zetterström, 2003), for example diffusion pathways in crystalline ionic conductors, but still based on structural data. This approach has then been extended (RMCt) to fit to the dynamical structure factor (Gereben, Pusztai, & McGreevy, 2007) by assigning a time step between configurations and thus creating an analogue of a MD simulation.

In EPSR an empirical potential is used in a Monte Carlo model to generate a structure and the potential parameters are refined based on the difference between experimental structural data and that calculated from the model. When the refinement has converged the potential can then be used on its own to create ensembles of structures. In principle the potential could also be used in a MD simulation to create dynamical models. However, it has been recognised for a long time that often a range of potentials can adequately reproduce the same set of structural data, whereas they might produce quite different results in terms of dynamics. Hence the use of the word 'empirical'; the potential is purely used as a tool for generating a structural model.

More recently, an approach to refining PE parameters from dynamical structure factor data has been reported in the detailed study by (Borreguero & Lynch, 2016). However, in this case a particular problem has been chosen where the data to be fitted are dominated by a single type of motion and hence a single PE parameter is refined, so it is not clear whether this algorithm can be generalised to a wider range of problems.

It might be questioned whether such model fitting to experimental data is any longer needed, since classical force fields/potentials and ab-initio methods can in many cases provide semi-quantitative agreement with experimental data. But the problem remains to ascertain how significant the deviations between simulation and experiment are for the purposes of understanding the relevant physical phenomena and, perhaps more importantly, how those deviations then affect the validity of any predictions that simulations may be used to make for either different conditions or different materials. Ab-initio methods also still tend to be restricted to smaller systems and timescales than would often be of interest.

The algorithm described in this paper takes strong inspiration from the RMC and EPSR methods to be generally applicable to any structural or dynamical data that can be calculated from a simulation, such as the dynamical structure factor. In practice the starting point would normally be the best available potentials.

The algorithm

The algorithm developed uses a standard classical MD simulation wrapped inside a Monte Carlo (MC) minimisation routine that optimises the potential energy parameters to provide the best fit between calculated and experimental data. An important development required to significantly decrease the computing time was to only use small changes of the parameters in order to maintain the simulation in equilibrium, i.e. any change in parameters should alter the total energy of the simulation by a smaller amount than the normal numerical fluctuation.

1. Specify the starting values of the PE parameters and run the initial MD simulation to equilibrium. Calculate the function(s) to be fitted (see below) and the Figure of Merit (FOM) function that measures the discrepancy between calculation and experiment(s).

2. Store the FOM value and MD phase-space configuration.

3. Select new values for the PE parameters by randomly modifying the previous PE parameters. Parameter changes should be sufficiently small to maintain the simulation in equilibrium. (At present the maximum parameter changes have been adjusted manually, however this could be automated.)

4. Run MD, re-calculate the FOM.

5. Use the Metropolis criterion to accept or reject the new FOM and therefore the PE parameters used to calculate the FOM, i.e. accept the new PE parameters with probability $\mathrm{P}_{\mathrm{accept}} = \min\left(1, exp\left(-\frac{FOM_{old}-FOM_{new}}{T_{MC}}\right)\right)$ where the MC 'temperature' T_{MC} controls the acceptance rate. If the MC step is accepted then overwrite the previously stored FOM value and phase-space configuration from step 2.

6. Repeat from step 3 until convergence has been achieved (the average value of the FOM over a number of iterations no longer decreases) or the chosen maximum number of iterations has been reached.

The FOM is defined as

$$FOM = \sum_i w_i \left[\frac{F_i^{data} - F_i^{cal}}{\sigma_i} \right]^2 \quad (1)$$

where F_i is any function that can be both determined from experimental data and calculated from the simulation, for example the dynamical structure factor S(Q, ω). σ_i is a measure of the experimental error which may be constant or a function of the same parameters as the data, e.g. Q and ω in the case of S(Q, ω). w_i is a weighting factor that assigns a relative importance to different data sets, for example accounting for the fact that they may have very different numbers of data points.

Although the algorithm has only been tested on a system with a single atomic species this is not a restriction.

Initial results

The algorithm has been tested using S(Q, ω) data for liquid argon collected at T=120K ρ=0.0176Å^{-3} (van Well et al., 1985) where the data are modelled with a Lennard-Jones (LJ) potential.

S(Q, ω) is related to the time-dependent correlation function G(r,t) through the Fourier transform

$$S(Q,\omega) = \frac{1}{2\pi} \int G(r,t) e^{i(Q \cdot r - \omega t)} dt dr \quad (2)$$

Ignoring Quantum effects, then in the classical limit of G(r,t) for a single atomic species is

$$G(r,t) = \frac{1}{N} \langle \sum_{i,j} \delta(r - r_j(t) + r_i(0)) \rangle \quad (3)$$

where <…> denotes a thermal average. This is calculated by averaging over statistically independent (or nearly independent) G(r,t) instances., which were determined from the MD simulations using the approach described in the book by Rapaport (Rapaport, 2004).

The partial differential cross section for inelastic neutron scattering from a single atomic species is proportional to $\sigma_{inc} S^s(Q,\omega) + \sigma_{coh} S(Q,\omega)$, where S^s is the Fourier transform of the self-part (i=j) of the correlation function in Eq. (3), σ_{inc} is the incoherent cross section and σ_{coh} is the coherent cross section (Squires, 1996). For argon the incoherent cross section can be neglected in which case the inelastic neutron scattering cross section is proportional to S(Q, ω). For a single atomic species the time-dependent correlation function and the space-time pair correlation function (as shown in Figure 1) are related by g(r,t)= G(r,t)/ ρ, where ρ is the atomic number density.

The MD simulation used a cubic box with standard periodic boundary conditions. The maximum values of r and t for the simulation need to be sufficient to avoid truncation of G(r,t) in the Fourier transform Eq.(2) that would otherwise introduce artefacts into the calculated S(Q, ω). In this case a

box containing 864 atoms was used. In addition the minimum number of time steps needs to allow for an equal number of time origins (Eq. (3)) for both the shortest and longest times. In this case the individual MD time step was 10.75 fs, and the number of time steps (after the initial run to equilibrium) was 3500. This time resolution was not required for the Fourier transform (determined by the maximum value of ω) so the G(r,t) histogram time bin was 7 MD time steps; the r bin was 0.2 Å.

The experiment above was repeated where the MD simulations were run with larger box sizes, different initial configurations of the atoms (simple cubic, face centred cubic and liquid) and for different times. It was found that for this simple system the algorithm could cope with this.

The experimental liquid argon data are shown in Figure 1(a). The initial values selected for the LJ parameters were ε=1.5K/mol and σ=4.0Å and, as can be seen from Figure 1(b), the system then equilibrates in a crystalline phase (face centred cubic). This marks the completion of steps 1 and 2 of the proposed algorithm.

The algorithm continues through cycles of updating LJ parameters; in this test it was allowed to run beyond convergence for 400 cycles/MC steps. The LJ parameters found with the smallest FOM were ε=1.020K/mol and σ=3.362Å and these may be compared to the published values: ε=1.0243K/mol and σ=3.36Å (van Well et al., 1985). Figures 1(c-d) show the corresponding $S^{cal}(Q, \omega)$ and space-time pair correlation function $g^d(r,t)$ for the best obtained LJ parameter values and, as expected, the MD simulation now equilibrates in the liquid state ($g^d = g - g^s$ is shown because otherwise the sharp spike in the bottom corner obscures the figure). Finally, figures 1(e-g) compare the experimental and calculated data for the best values of the LJ parameters.

The calculated FOM values (on a log scale) as a function of MC steps (cycles) are shown in Figure 2, where the crosses plot the accepted MC steps and the circles plot the rejected MC steps. Within the first 20 MC steps the FOM value drops sharply transforming the system from a crystalline state to a liquid state and the system thereafter stays in this state. Convergence appears to have been reached after 200 steps. It was then tested whether lowering the T_{MC} parameter from T_{MC}=1 to T_{MC}=0.25 would further reduce the FOM; the percentage of accepted steps dropped from ~32% to ~18% but otherwise no noticeable change was observed.

Figure 3 show the pairs of values of σ and ε for all the accepted MC steps starting from the 15th accepted step, with the '+' in the top right-hand corner showing the starting parameters. The point where the horizontal and vertical dashed lines cross marks the published values for liquid argon in (van Well et al., 1985). The (σ,ε) pair of values which is closest happens to be that with the smallest FOM value. However, this should be viewed as coincidental since fluctuations of order $\pm$ 10% in ε and $\pm$ 3% in σ would produce equally acceptable fits to the experimental data. The distribution of value pairs also indicates some correlation, with higher values of ε corresponding to lower values of σ and vice-versa.

Figure 4 shows the values of the diffusion constant for the accepted MC steps. After about 20 MC steps the system settles into the liquid state with a calculated diffusion constant fluctuating near the published value for this system: D=0.68 10^{-8} m^2s^{-1} (van Well et al., 1985).

To reproduce these results the MDMC software can be downloaded from [software DOI reference]. Note that because of the random moves of the PE parameters and initial scaled velocities of the starting phase-space configuration no two runs of MDMC will yield exactly the output.

Conclusions

A new algorithm has been proposed to model dynamical structure factor data from inelastic neutron scattering, which was heavily inspired by the RMC and EPSR methods. As for these algorithms, the proposed method may also be easily combined with other types of data and constraints. Initial results from using the algorithm were presented, which showed that it was able to straightforwardly optimise LJ parameters from a poor initial starting point of these parameters.

Acknowledgements

Thanks to Keith Refson for advice on how to get started with writing an MD engine and for pointing out the book by D. C. Rapaport. Also, thanks to Jon Wakelin for answering questions related to the XML Fortran library xmlf90.

References

Borreguero, J. M., & Lynch, V. E. (2016). Molecular Dynamics Force-Field Refinement against Quasi-Elastic Neutron Scattering Data. *Journal of Chemical Theory and Computation*, *12*(1), 9–17. https://doi.org/10.1021/acs.jctc.5b00878

Bouty, O. (2014). Application of the Empirical Potential Structure Refinement Technique to a Borosilicate Glass of Nuclear Interest. *Procedia Materials Science*, *7*, 32–37. https://doi.org/10.1016/j.mspro.2014.10.006

Gereben, O., Pusztai, L., & McGreevy, R. L. (2007). Development of the time-dependent reverse Monte Carlo simulation, RMC *t*. *Journal of Physics: Condensed Matter*, *19*(33), 335223. https://doi.org/10.1088/0953-8984/19/33/335223

McGreevy, R. L. (2001). Reverse Monte Carlo modelling. *Journal of Physics: Condensed Matter*, *13*(46), R877–R913. https://doi.org/10.1088/0953-8984/13/46/201

McGreevy, R.L. and Zetterström, P. (2003) To RMC or not to RMC? The use of reverse Monte Carlo modelling. *Current Opinion in Solid State and Materials Science, 7*, 41-47

McGreevy, R. L., & Pusztai, L. (1988). Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures. *Molecular Simulation*, *1*(6), 359–367. https://doi.org/10.1080/08927028808080958

Playford, H. Y., Owen, L. R., Levin, I., & Tucker, M. G. (2014). New Insights into Complex Materials Using Reverse Monte Carlo Modeling. *Annual Review of Materials Research*, *44*(1), 429–449. https://doi.org/10.1146/annurev-matsci-071312-121712

Rapaport, D. C. (2004). *The art of molecular dynamics simulation*. Cambridge University Press.

Soper, A. K. (1996). Empirical potential Monte Carlo simulation of fluid structure, *202*, 295–306.

Soper, A. K. (2007). On the uniqueness of structure extracted from diffraction experiments on liquids and glasses. *Journal of Physics: Condensed Matter*, *19*(41), 415108. https://doi.org/10.1088/0953-8984/19/41/415108

Soper, A. K., & Edler, K. J. (2017). Coarse-grained empirical potential structure refinement: Application to a reverse aqueous micelle. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1861*(6), 1652–1660. https://doi.org/10.1016/j.bbagen.2017.02.028

Squires, G. L. (1996). Introduction to the Theory of Thermal Neutron Scattering. Dover Publications. https://doi.org/10.1017/CBO9781139107808

van Well, A. A., Verkerk, P., de Graaf, L. A., Suck, J.-B., & Copley, J. R. D. (1985). Density fluctuations in liquid argon: Coherent dynamic structure factor along the 120-K isotherm obtained by neutron scattering. *Physical Review A*, *31*(5), 3391–3414. https://doi.org/10.1103/PhysRevA.31.3391

Captions

Figure 1: (a) Shows the T=120K, ρ=0.0176Å^{-3} argon S(Q, ω) data from (van Well et al., 1985) multiplied by $\exp(\hbar\omega/(2k_BT))$. This multiplication is performed to 'un-symmetrise' the symmetrised S(Q, ω) data, see Eq. (5) in (van Well et al., 1985). (b) Shows the distinct part of the space-time pair-distribution function, $g^d(r,t)$, for the first guess at the PE parameter values. (c-d) Show plots of $S^{cal}(Q, \omega)$ and $g^d(r,t)$ for the PE parameter values resulting in the smallest FOM value found. (e) Shows the difference between S(Q, ω) in (a) and (c). The solid line in (f) shows $\int S^{data}(Q,\omega)d\omega$ and the dashed line $\int S^{cal}(Q,\omega)d\omega$. S(Q, ω) is in units of 10^{-13}s = 0.1 ps. $g^d(r,t)$ is dimensionless.

Figure 2: Shows the FOM value, on a log scale, for accepted (crosses) and rejected (circles) MC steps. The FOM here is as defined in Eq. (1) with the weighting factors set to one and F_i set to the dynamical structure factor. During the first 200 MC steps T_{MC}=1. The MDMC simulation was then continued, after the first 200 MC steps, using as a starting point the LJ parameters with the lowest FOM value from the first 200 MC steps and with T_{MC} reduced by a factor of four.

Figure 3: Plots the LJ parameter values of the accepted MC steps as crosses after the 15th step. The '+' at the top right-hand corner shows the values of the LJ parameters first selected. The horizontal and vertical dashed lines show the published (van Well et al., 1985) LJ parameter values: ε=1.0243K/mol and σ=3.36Å.

Figure 4: Shows the value of the calculated diffusion constant in units of [10^{-8} m^{2}s^{-1}] for all accepted MC steps. The horizontal solid line plots the published value for the diffusion constant: D=0.68 10^{-8} m^{2}s^{-1} (van Well et al., 1985).

Figure 1

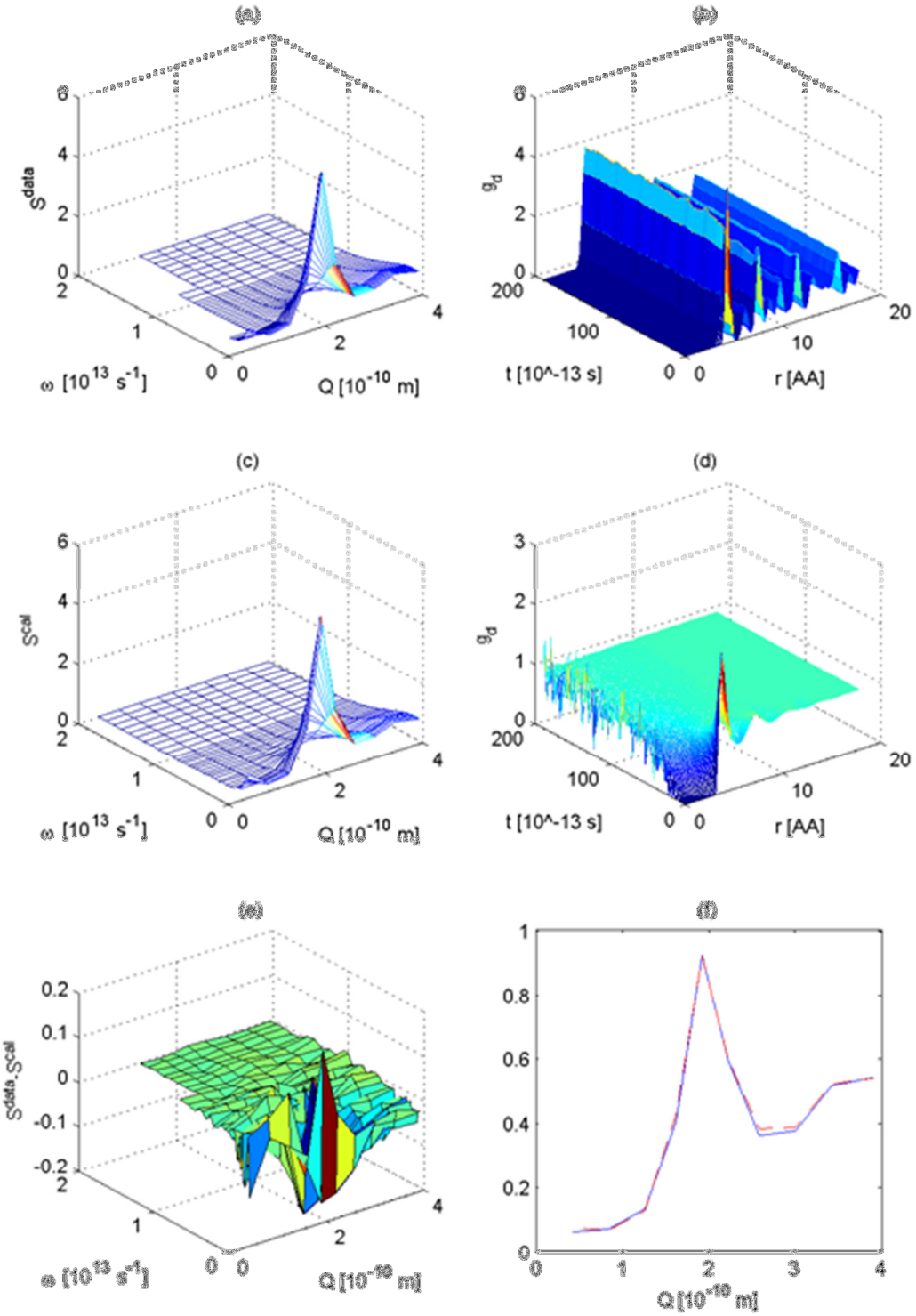

(a)
S^data
ω [10^13 s^-1]
Q [10^-10 m]
(b)
g_d
t [10^-13 s]
r [AA]
(c)
S^cal
ω [10^13 s^-1]
Q [10^-10 m]
(d)
g_d
t [10^-13 s]
r [AA]
(e)
S^data-S^cal
ω [10^13 s^-1]
Q [10^-10 m]
(f)
Q [10^-10 m]

Figure 2

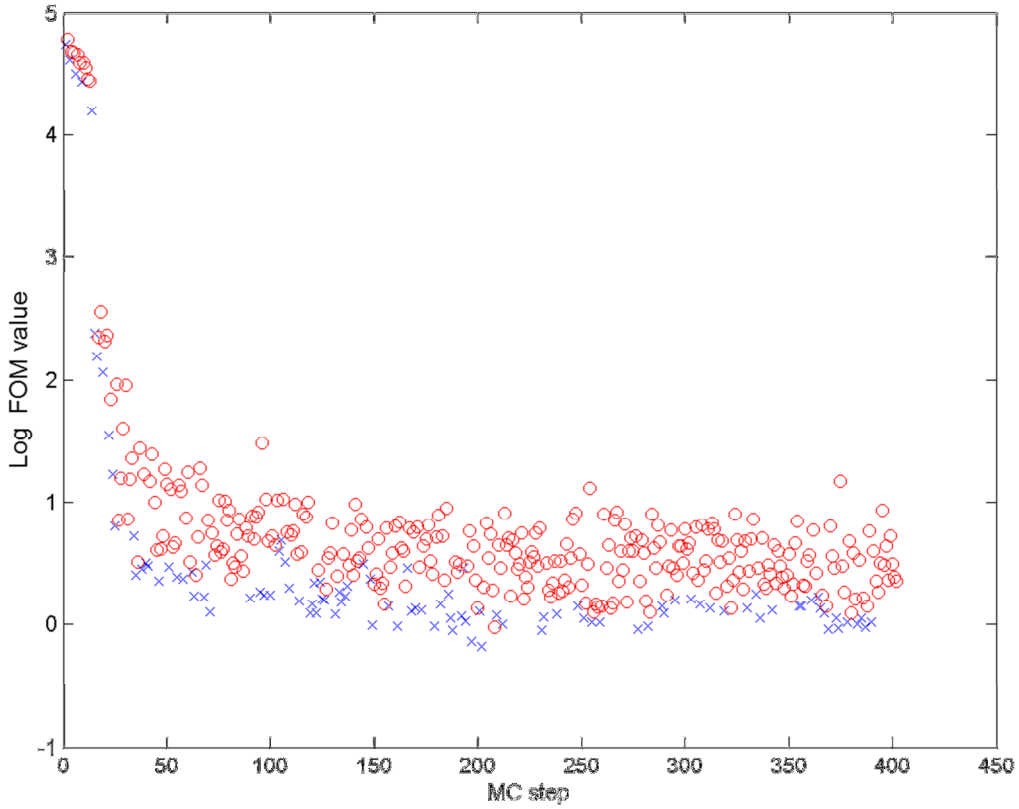


Figure 3

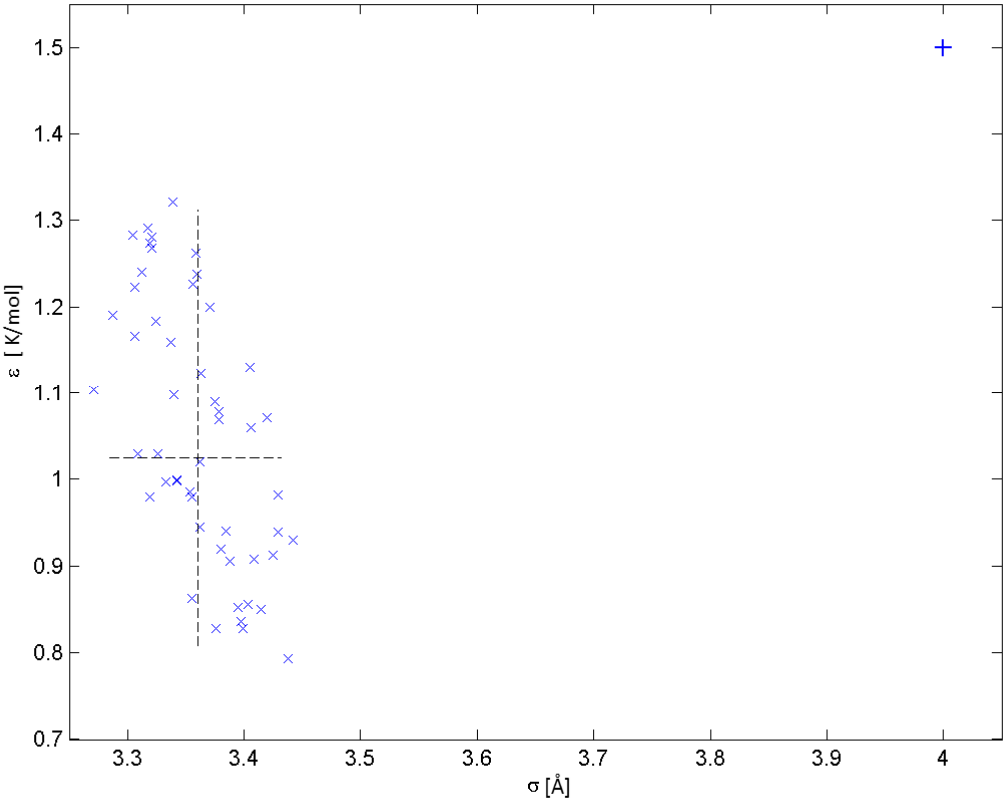
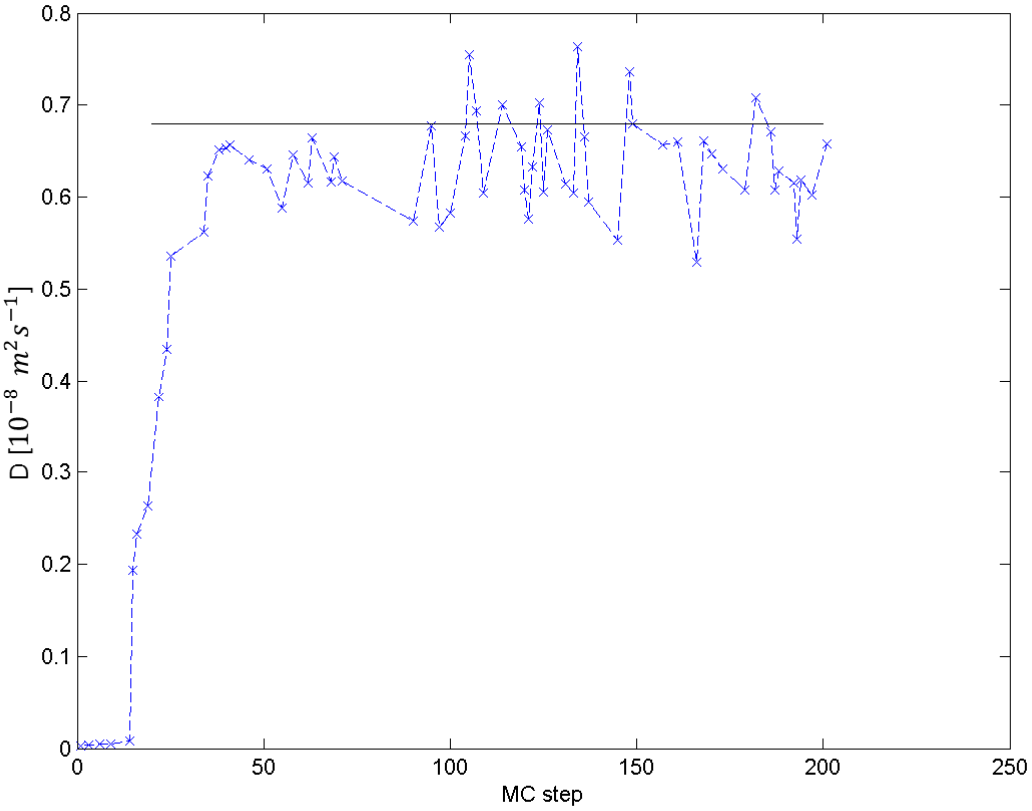


Figure 4


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
D [10^-8 m^2 s^-1]
0
50
100
150
200
250
MC step