Scientific Report for ESF short visit grant 2296

Adam Swetnam
Department of Physics, University of Warwick, Coventry, CV4 7AL, UK

I. PROJECT DESCRIPTION

This project is concerned with the Monte Carlo (MC) simulation of simple lattice models of peptides. Their behaviour near adsorbing surfaces is of particular interest due to the technological possibilities relating to the design of surfaces that are highly selective for particular biological molecules [1]. Although very simplistic, lattice models capture the basic aspects of the problem including the difficulty of finding minimum-energy folded structures.

In the HP lattice model [2] a peptide is represented by a self-avoiding chain of hydrophobic (H) and polar (P) beads residing on a lattice. Each non-bonded pair of H beads occupying neighbouring lattice sites contributes an energy $-\varepsilon$. Homopolymers may be represented by the interacting self-avoiding walk which is equivalent to the HP model with all H beads. In order to study the behaviour of the model it is necessary to sample effectively from low-energy states which is difficult due to the rough energy landscape of the model.

One approach to sampling effectively in systems with rough energy landscapes is density-of-states sampling. Density-of-states methods attempt to sample equally across the entire energy range of a system, and in order to do this they must be able to generate $g(E)$, an estimate for the density of states of the system as a function of energy E. The Wang-Landau algorithm [3, 4] is an efficient, generic density-of-states method.

In The Wang-Landau algorithm a move selected from a move set is performed to generate state B with energy E_B from original state A, energy E_A and, assuming an unbiased move set is used, this move is accepted with probability

$$P_{\mathrm{acc}}(A \to B) = \min\left(1, \frac{g(E_A)}{g(E_B)}\right) . \tag{1}$$

After each move $g(E_C) \to f \times g(E_C)$ where C is the current state of the system and f is the 'modification factor' which asymptotically approaches unity as the simulation progresses. The result is that $g(E)$ converges to some extent towards the true density of states and states are sampled relatively equally across the entire range of energies.

In the original formulation of the Wang-Landau algorithm a histogram $H(E)$ of the number of visits to each energy is kept. Periodically it is checked whether the histogram meets the flatness criterion and if $\min(H(E)) > \phi \times \overline{H}$, where $\overline{H}$ is the mean over all available energies of the system and $0 < \phi < 1$ a parameter of the simulation, then $f \to \sqrt{f}$ and the histogram is reset. Using this method $g(E)$ initially rapidly approaches the true density of states; however the convergence eventually stops and an alternate algorithm has been suggested [5, 6]. In the $1/t$ Wang-Landau algorithm the original method is followed initially, but at long times t, in order for convergence to continue, $\ln f$ is set equal to $1/t$. Here $t = M/N_E$ where M is the number of MC moves attempted so far in the simulation and N_E the number of energies available.

In order for the Wang-Landau algorithm to access all low-energy states it must be used with a move set which is ergodic and in order for $g(E)$ to converge to the correct value the move set must also be reversible. Pull moves [7] can be proven to be ergodic and reversible but the method for selecting pull moves is *biased* so that it is necessary to count all possible forward and reverse moves each time one was attempted. Calculating all possible pull moves each step is quite time consuming, and can dominate the calculation for long chains.

A peptide near an adsorbing surface can be respresented by labelling a plane of lattice sites as a wall. Each bead occupying a site adjacent to the wall contributes an energy ε_{s}. The density of states may now be defined as a function of two variables, the bead-bead interaction energy E_{b}, and the bead-surface interaction energy E_{s}: $g(E_{\mathrm{b}}, E_{\mathrm{s}})$ [8]. In order that all elements of $g(E_{\mathrm{b}}, E_{\mathrm{s}})$ are finite a, second, confining but non-interacting wall must be introduced so that there are L lattice planes between the two walls. $g(E_{\mathrm{b}}, E_{\mathrm{s}})$ can be used to calculate heat capacity as a function of two variables and thus to plot pseudo-phase diagrams as a function of temperature and solvent quality.

II. VISIT DESCRIPTION

The purpose of the visit was to discuss the application of Monte Carlo techniques to the study of the adsorbtion of coarse-grained models of peptides on inorganic surfaces. During the short visit I had discussions with several researchers:

Prof. Kurt Binder Application of the Wang-Landau algorithm to polymer molecules near surfaces. Specifically the quantities which can be used to find and characterise phase transitions and move sets for lattice polymers were discussed.

Dr. Hsiao-Ping Hsu The Pruned-Enriched Rosenbluth Method (PERM) and its efficiency for HP lattice models.

Prof. David Landau Technical details of the Wang-Landau algorithm including its convergence and the importance of finding low energy states and also move sets that can be employed to speed up sampling of compact conformations of lattice polymers.

Prof. Wolfgang Paul Technical details of the Wang-Landau algorithm as applied to polymer molecules. In particular how to deal with the fact that the ground state energies are not known *a priori* and how to characterise the phase transitions of polymers near surfaces.

Dr. Nico van der Vegt Relating the idealised model used in this project to more chemically realistic models of peptides near surfaces and the phenomenon of surface specificity of peptides.

Additionally as the dates of my visit coincided with the University of Mainz MATCOR "Simulation of Macromolecules on Different Scales" summer school I took the opportunity to attend and, in order to take advantage of the grant awarded, extended my stay in Mainz accordingly. The school consisted of seminars on topics closely related to the aims of the SimBioMa program and also a Poster session at which I was able to present my work to date.

The visit has aided the development of the project by allowing interaction with researchers at the host institution. The complementary nature of the research being carried out by the two groups has been highlighted, so no collaborative projects have been initiated and no joint publications are expected. Work for the project completed before the visit has now been published [9] and further publications related to the project are expected.

III. PROJECT PROGRESS

For this project an unbiased method of selecting pull moves has been developed [9]. If chain overlaps are ignored in generating pull moves the number of moves for all conformations is equal so that it is not necessary to count them each time a move is attempted. It is still necessary to select such moves with equal probability and not all beads have an equal number pull moves available to them. However, the number of available pull moves (ignoring chain overlap) for a given bead does not depend on the specific chain conformation, but simply on the location of the bead in the chain so selection of these pull moves with equal probability is easily and quickly done. Of course, trial conformations generated this way may contain overlaps, so more moves will be rejected. A compromise between biased and unbiased selection was also developed in which some counting of moves was necessary but the number of rejected moves was reduced.

Fig. 1 shows the time taken per MC move for various length lattice homopolymers, run lengths of 10^5 moves were used. Only a certain fraction of moves are acceptable for the unbiased and compromise moves and this fraction is shown in the inset, on a log-log plot. This fraction is minimised in the compact globule phase and for unbiased pull moves scales roughly as $N^{1/3}$, reflecting the fact that most acceptable moves involve rearrangement at the compact globule surface. Move times corrected for this factor are also given for the biased and compromise move sets. Over this range of chain lengths, the simpler unbiased selection procedure, even with its limited move acceptance rate, is more than an order of magnitude faster. The compromise pull moves are faster than traditional pull moves, but slower by about a factor two than the unbiased moves. However, because their acceptance ratio does not scale so strongly with chain length, they may become competitive for longer chains.

A new method of simulating lattice peptides near adsorbing surfaces has also been devised [9]. Instead of labelling a certain lattice plane as the interacting surface, the lattice plane adjacent to the peptide is always labelled as the surface. In this way the density of states for a lattice peptide confined to be on the surface can be determined. It can be shown that by determining the average of an additional quantity (the height of conformations) the density of states for the slit system can be exactly determined from this new density of states. However, the new method is more efficient than the slit method because:

1. simulation of essentially equivalent peptide configurations detached from the wall is avoided;

2. no moves are rejected because they generate conformations which intersect the wall.

Fig. 2 shows the number of unbiased pull moves required for convergence of Wang-Landau simulations of various length homopolymers for both systems. In each case, for slit geometry, the minimum possible wall separation $L = N$ was used. The quantity measured was the time taken to complete the 16th iteration, i.e. to reach a final modification

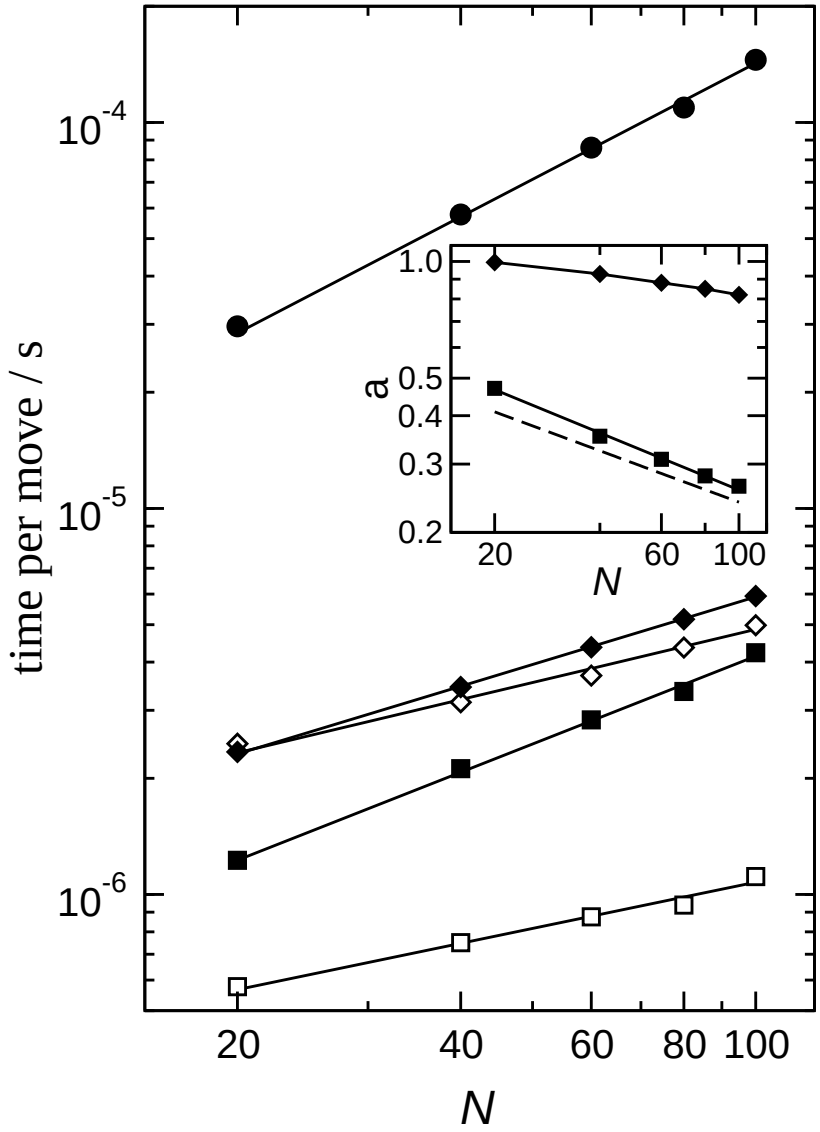

FIG. 1: CPU times per move plotted against chain length N, both on logarithmic scales, for self-avoiding lattice polymer chains in the compact globule phase. Circles, biased pull moves; squares, unbiased pull moves; diamonds, compromise pull moves. Open symbols give uncorrected rates, filled symbols give rates corrected by fraction of acceptable moves. The inset shows the fraction of moves which are acceptable (generate non-overlapping conformations) as a function of chain length, again both on logarithmic scales, for unbiased pull moves (squares) and compromise pull moves (diamonds); the dashed line has gradient $-1/3$ for comparison. In all cases, statistical errors are smaller than the symbol sizes.

factor f given by $\ln f = 1/2^{16} \approx 1.53 \times 10^{-5}$. Normally, the runs would be continued well beyond this point, but the variation in completion times becomes quite significant in the latter stages. The results were averaged over 10 independent Wang-Landau runs. The figure shows that the new method of sampling is about a factor of 3 quicker and it must be noted that the new method provides the slit system density of states for all L not just a single value.

Also in the project the $1/t$ Wang-Landau algorithm has been applied to lattice polymers for the first time and a comparison with the original method performed. The $1/t$ Wang-Landau method assumes *a priori* knowledge of N_E which is not available for lattice peptides, so N_E must be taken to be the number of energies so far found to be possible. Fig. 3 shows a measure of the error in the estimate of the density of states against Monte Carlo moves attempted for a short 18-monomer homopolymer for which the exact density of states can be determined by enumeration. 10 runs were performed for each data set and the error bars represent the standard error of the mean. It can be seen that the error for the $1/t$ method is always less than or equal to (within uncertainty) the error in the original method with various ϕ.

In conclusion three improvements to the method for simulating lattice peptides near adsorbing surfaces have been

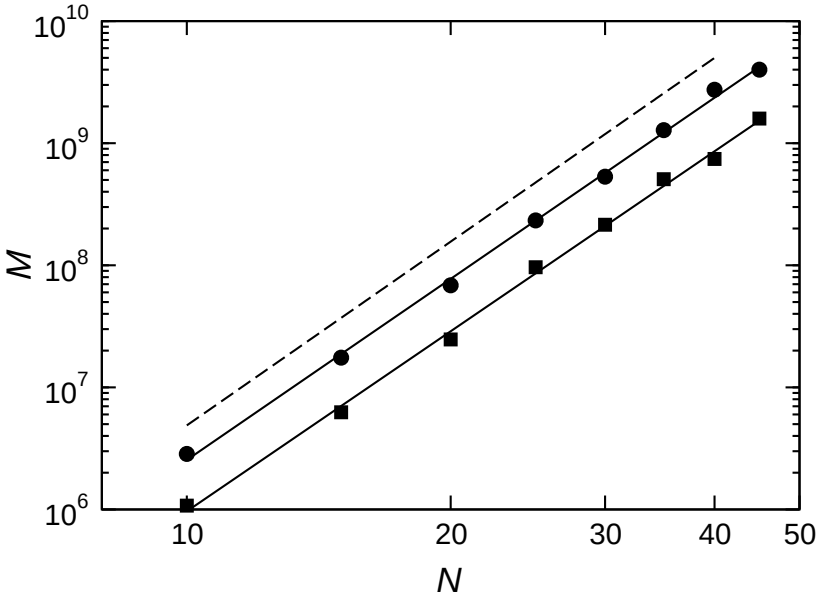

FIG. 2: Number of Monte Carlo unbiased pull moves M to convergence of Wang-Landau simulations of polymer chains, for various chain lengths, plotted on log-log scales. Circles, slit geometry with slit width $L = N$; squares, conformations confined to the surface. The dashed line has gradient 5, for comparison. In all cases, statistical errors are smaller than the symbol sizes.

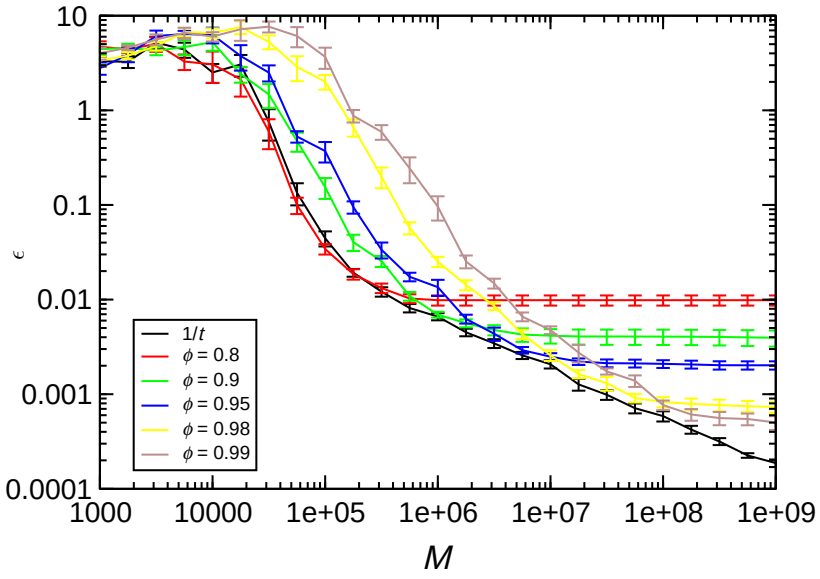

FIG. 3: Error ϵ against MC moves for an 18-momomer homopolymer using the original Wang-Landau algorithm with various ϕ and the Wang-Landau $1/t$ method with $\phi = 0.8$.

made:

1. An unbiased method of selecting pull moves which proves more efficient has been developed;
2. A method of simulating a peptide confined to be in contact with a surface has been devised, and shown to allow results for slit geometry with any L to be calculated, despite taking less simulation time than a slit geometry simulation for a single L;
3. The $1/t$ Wang-Landau algorithm has been applied to lattice polymers for the first time and been shown to be an improvement on the original algorithm.

These advances in techniques will be used to study the adsorption of lattice petides and determine what can be learnt

about surface specificity from these simple models.

[1] M. Sarikaya, C. Tamerler, D. T. Schwartz, and F. O. Baneyx, *Ann. Rev. Mater. Res.*, 2004, **34**, 373–408.
[2] K. F. Lau and K. A. Dill, *Macromolecules*, 1989, **22**, 3986–3997.
[3] F. G. Wang and D. P. Landau, *Phys. Rev. Lett.*, 2001, **86**, 2050–2053.
[4] F. G. Wang and D. P. Landau, *Phys. Rev. E*, 2001, **64**, 056101.
[5] R. E. Belardinelli and V. D. Pereyra, *Phys. Rev. E*, 2007, **75**, 046701.
[6] R. E. Belardinelli and V. D. Pereyra, *J. Chem. Phys.* , 2007, **127**, 184105.
[7] N. Lesh, M. Mitzenmacher, and S. Whitesides In ed. M. Vingron, S. Istrail, P. Pevzner, and M. Waterman, *Proceedings of the 7th Annual International Conference on Research in Computational Molecular Biology*, RECOMB, pp. 188–195, New York, 2003. Association for Computing Machinery.
[8] M. Bachmann and W. Janke, *Phys. Rev. Lett.*, 2005, **95**, 058102/1–4.
[9] A. D. Swetnam and M. P. Allen, *Phys. Chem. Chem. Phys.* , 2009, **11**, 2046–2055.