A Parallel Cluster Algorithm Applied to Model DNA Systems

J. Schluttig[1,2] and G. Sutmann[2]

[1] Institut für Theoretische Physik, Universität Leipzig, Germany

[2] Zentralinstitut für Angewandte Mathematik, John von Neumann Institut für Computing, Forschungszentrum Jülich, Germany

Abstract

The aim of this work [1] is to build the basis for a parallel Monte Carlo cluster algorithm for continuous two-dimensional spin systems. A purely geometrical technique for searching clusters is developed. It is capable of being extended to an energetic cluster criterion, which is the basis in Monte Carlo cluster methods. The scaling of the implementation is measured and analyzed. The new method is applied to model DNA systems [2] simulated with *"SpinCG^{2d}"* [3] using the Kornyshev-Leikin [4] potential. Geometric clusters are studied as a function of different DNA characteristics, e.g. the charge compensation parameter θ.

DNA Model System

It is well known that DNA forms close-packed aggregates of various structures, e.g. in human chromosomes or viruses. Experimentally it was observed that short fragments form columnar aggregates which are suitable to study interactions [5].

A pair potential for the interaction of two parallel DNA fragments of persistence length L_p can be derived [4], describing them as long cylinders, carrying helical, continuous line charges on their surface. The only dependencies are the distance $r = |\mathbf{R}|$ and the azimuthal orientation $\phi = \phi_1' - \phi_2'$ (see fig. 1). Thus the model can be treated as a continous two-dimensional spin system.

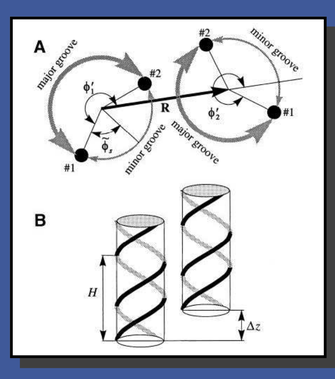
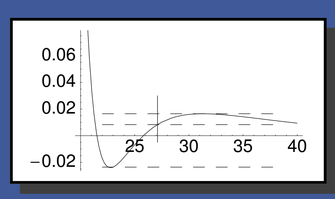


Figure 1: Important structural values for two interacting DNA double strands.

Cluster Criterion

It is necessary to introduce an adequate threshold distance r_0, which decides whether two molecules belong to the same "geometric cluster". It turned out that correlating r_0 with the potential as the distance, where the potential barrier is overcome by $4/5$, provides acceptable results.



The Algorithm

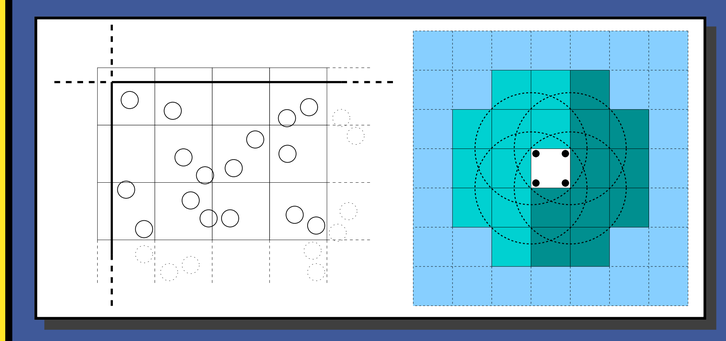
The cluster search has to deal with the continuous properties of the system and the particular parallelization of *"SpinCG^{2d}"*, utilizing a spatial decomposition and being capable of running on distributed memory systems.

The basic approach consists of:

1. identifying clusters sequentially and independently on each node,
2. communicating to link global clusters,
3. distributing the whole linking information.

Local Cluster Identification

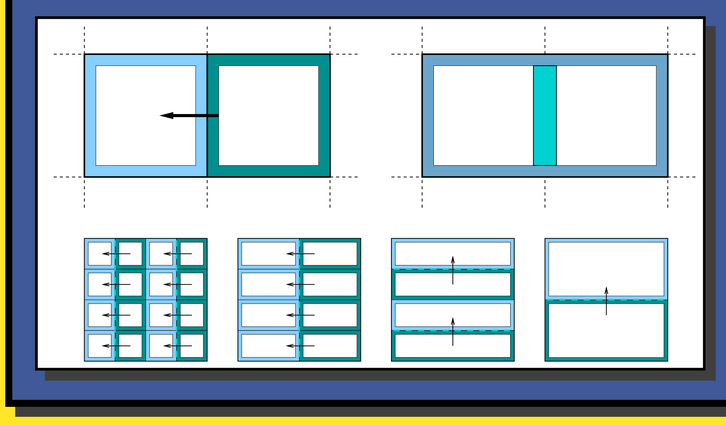
Concerning the cluster identification a lattice system has two advantages compared to a continuous system: The particles are in a given order and there is a definite number of neighbors. Thus the idea is to extend the highly efficient but lattice-based Hoshen-Kopelman [6] algorithm.



The figure shows how the molecules are sorted into a virtual square lattice structure at the border of a node. The lattice constant is chosen as $l = r_0/\sqrt{2}$ so that two molecules belonging to the same lattice cell are geometrically bound.

Global Cluster Linking

For the global linking some kind of *all-to-all* communication is needed. The figure below illustrates the chosen communication model. One of two communicating nodes is the *master* and receives the whole domain edge information of the *slave*. Now the bordering edge is linked and the communication is continued with a master of the same rank. The amount of linking data cannot be reduced, but in so doing the number of communication steps scales with $\mathcal{O}(\ln N_p)$.



Scaling

The theoretical speedup for a system with latency λ and throughput χ is:

$$c(N_p) = \log_2(N_p)\lambda + \log_2(N_p)\left(1 - \frac{1}{N_p}\right)\chi \ .$$

Fig. 2 clarifies that $c(N_p)$ can imply negative speedups on some algorithmic part, which however does not have a crucial influence on the entire scaling as the actual runtime is noticeable below the Monte Carlo routines.

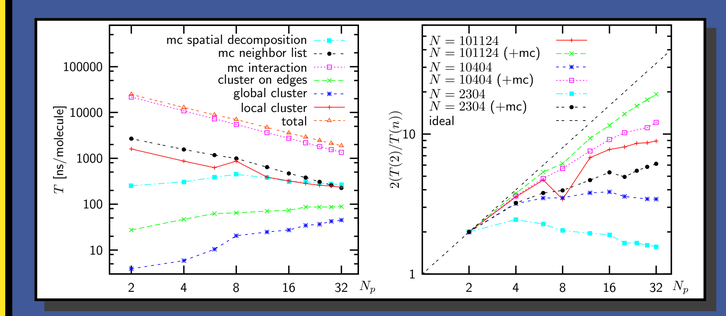


Figure 2: (a) Runtimes of different routines. (b) Scaling of the cluster algorithm itself and attached to the Monte Carlo step.

Simulations

All structural constants are chosen as appearing in B-DNA. Groundstate-like structures are obtained by cooling down a system of 5000 particles linearly from 1000K to 30K in varying numbers of Monte Carlo Metropolis steps, starting from a hexagonal structure with a lattice constant of $d = 35.0$Å. The system structure is analyzed by measuring cluster size histograms. The degree of compensation of the negative charged phosphates by adsorbed counter ions, parameterized by θ, is varied. The choice of θ is crucial for the interaction.

Cluster Size Distributions

The errorbars on the cluster size histograms result from averaging over 100 independent final configurations. From percolation theory it is known that the relative number n_s of clusters of size s is asymptotically given by

$$n_s \sim s^{-\tau} e^{-cs} \ .$$

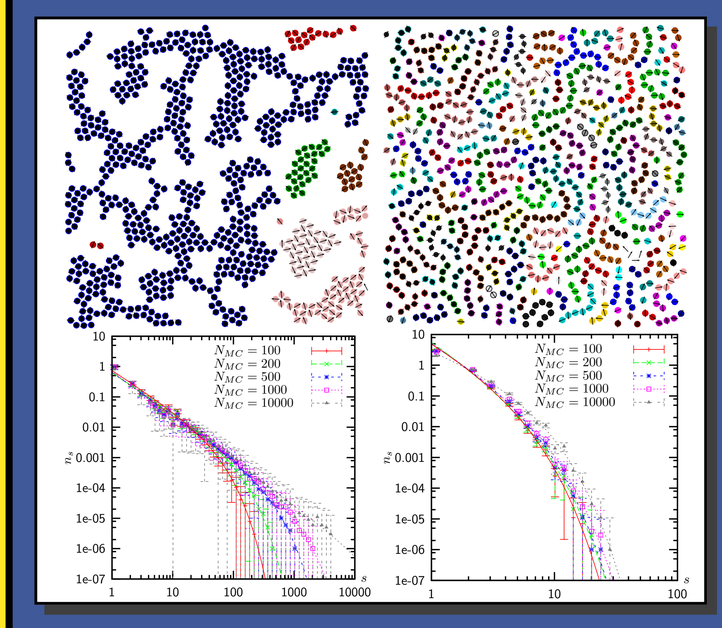


Figure 3: (a), (b) Section of a final configuration after 10000 Monte Carlo steps for $\theta = 0.95$ and $\theta = 0.865$. (c), (d) The respective cluster size histograms with n_s fits.

In analogy to p_c in percolation studies, some critical value θ_c can be introduced for the observed system. In fig. 4 it can be seen, that c obviously diverges with a potential law close to θ_c. The critical exponent seems to depend on the speed of cooling down the system.

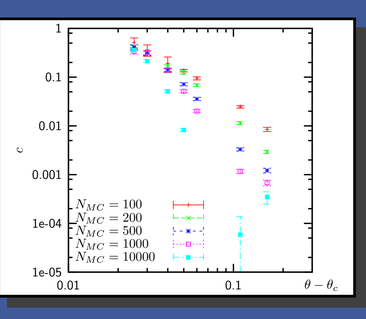


Figure 4: Double-log plot of the exponential factor c versus charge compensation.

References

[1] J. Schluttig, in *Technical Report IB-2004-11*, R. Esser (Ed.), John von Neumann Institute for Computing, Jülich, 2004

[2] H. M. Harreis, A. A. Kornyshev, C. N. Likos, H. Löwen, G. Sutmann, Phys. Rev. Lett. **89**, 018303 (2002)

[3] G. Sutmann, *SpinCG^{2d} - a parallel Monte Carlo program for spin systems*, in preparation

[4] A. A. Kornyshev, S. Leikin, J. Chem. Phys. **107**, 3656 (1997)

[5] D. C. Rau, V. A. Parsegian, Biophys. J. **61**, 246 (1992)

[6] J. Hoshen, R. Kopelman, Phys. Rev. B **14**, 3438 (1976)

Gedruckt am Rechenzentrum der Universität Leipzig