

Lattice Molecular Automaton (LMA): A Physico-Chemical Simulation System for Constructive Molecular Dynamics

Bernd Mayer
Steen Rasmussen

SFI WORKING PAPER: 1996-08-059

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

1 Authors to whom correspondence should be addressed.

e-mail: steen@lanl.gov

USA

Los Alamos, New Mexico, 87545

Los Alamos National Laboratory

•TSA-DO/SA MS-M997 and CNS MS-B258

USA

Santa Fe, New Mexico, 87501

1399 Hyde Park Road

^bSanta Fe Institute

e-mail: bernd@astrix.msp.univie.ac.at

Austria

Altthamistrasse 14, A-1090 Wien

Universität Wien, UZAII

^aInstitut für Theoretische Chemie und Strahlenchemie

Bernd MAYER,^a and Steen RASMUSSEN^{b,c}

Lattice Molecular Automation (LMA): A
Physico-chemical simulation system for
constructive molecular dynamics.

Molecular self-assembly is frequently encountered in biochemical systems generating higher order structures with well defined functionalities. However, the driving forces underlying these processes are not well understood. The Lattice Molecular Automation (LMA) is a computational tool suitable for simulating self-organization processes in large scale, molecular automation environments. This paper introduces the basic computational concepts needed to formulate molecular dynamics and self-assembly in a discrete field, cellular automation environment: Molecular objects are encoded as data structures on a hexagonal lattice. Propagating force particles together with kinetic and potential energy terms define simulation objects with a minimum complexity (number of physical variables together with interaction functions) with respect to specified molecular dynamics and force field properties. In this paper we focus on the mathematical and algorithmic formulation of a variety of intermolecular interactions through a decomposition of molecular type-specific force fields into ulnar interactions. As an example, the simulation of polymer dynamics in an aqueous environment is shown. The straight forward implementability of the LMA concept on massively parallel architectures as well as possible applications in the field of Computational Nanotechnology are briefly discussed. Thermodynamical properties together with a variety of other physico-chemical properties of the LMA are discussed in detail in reference [1].

Abstract

processes based on Newton's equations of motion [9], these interactions are modeled encoded in the molecular objects. In the common molecular dynamics simulation application rules with other molecules as the formation of hydrogen bonds, have to be interaction rules is the hydrophobic tail and a hydrophilic head. Such properties, or the concomitant (i.e. other molecules of various types). Intermistic features of e.g. amphiphilic molecules (i.e. encoding only the relevant interactions between the molecules and their environment that the molecular system is able to generate certain observables. This is achieved by chemical information at the molecular (object) level that is necessary and sufficient so the case of molecular systems, the minimum object complexity defines the amount of higher order structures or other emergent properties generated by the dynamics. In needs to be encoded on an object sufficient for the generation of particular, specified portant. The critical object complexity defines the least amount of information that In a theoretical context the notion of critical object complexity [4, 5] becomes important. of the single molecules) remains unclear.

[7, 8], but the cause for this sensitivity on the molecular scale (i.e. the microdynamics structures as e.g. vesicles on the chemical properties of the underlying lipid molecules experimental work shows a sensitive dependence of the morphology of supramolecular dynamic phenomena in molecular systems is barely treated.

Whereas the elements of a conceptual framework as well as some of the formal dynamical properties for the emergence of higher order structures in formal systems was described recently [4, 5, 6], the physico-chemical background for the generation of such processes is: What are the driving forces for the formation of such highly ordered structures – and what is the underlying chemical information exactly necessary for

processes is: An extensively studied example of molecular self assembly is vesicle and membrane formation [2, 3], both composed of amphiphilic polymers solvated in the biological solvent water. The basic question to ask about this and other defined functionalities. An extensively studied example of molecular self assembly well erly of biochemical systems - and in general macromolecular ensembles exhibiting well formation of hierarchically organized structures seems to be a fundamental property of biochemical systems - and in general macromolecular ensembles exhibiting well

terms

1.1 Emergence of Hierarchical Structures in Biochemical Sys-

1 Introduction

strict manner.

The LMA molecular dynamics is defined by two reservoirs of energy, the kinetic and potential energy. The kinetic energy is formally driving the molecular system into low energy configurations on the potential energy hypersurface. Kinetic energies are distributed over the (in the case of a hexagonal lattice three) degrees of freedom, following e.g. a Boltzmann distribution, and are distributed to other molecular objects in a collision process applying a hard sphere model. Each molecule has an excluded volume and pairwise collisions conserve momentum in a strict sense. Higher order free molecule collisions and collisions involving polymers do not conserve momentum in a volume and pairwise collisions conserve momentum in a strict sense.

The LMA molecular dynamics is formally implemented in a cellular automation simulation environment, discretizing space, time, matter and forces. Integer operations are used exclusively throughout the computation. The physical space is a two dimensional, hexagonal lattice, where the molecular objects occupy the sites of the lattice. The time is defined as synchronically performed update steps and is driven by an external clock.

The LMA molecular dynamics is formally implemented in a cellular automation simulation environment, discretizing space, time, matter and forces. Integer operations are used exclusively throughout the computation. The physical space is a two dimensional, hexagonal lattice, where the molecular objects occupy the sites of the lattice. The time is defined as synchronically performed update steps and is driven by an external clock.

The LMA molecular dynamics is defined by two reservoirs of energy, the kinetic and potential energy. The kinetic energy is formally driving the molecular system into low energy configurations on the potential energy hypersurface. Kinetic energies are distributed over the (in the case of a hexagonal lattice three) degrees of freedom, following e.g. a Boltzmann distribution, and are distributed to other molecular objects in a volume and pairwise collisions involving polymers do not conserve momentum in a strict sense.

1.2 Discrete Representation of Molecular Systems

The Lattice Molecular Automation is designed in the spirit of the Lattice Gas Automata [11] and is an extension of the Lattice Polymer Automata [12], which both proved to be capable of generating macroscopic effects only based on a discrete, microscopic system representation. The interaction rules in the LMA emphasize especially a correct treatment of interactions at the atomic scale.

The Lattice Molecular Automation is designed in the spirit of the Lattice Gas Automata [11] and is an extension of the Lattice Polymer Automata [12], which both proved to be capable of generating macroscopic effects only based on a discrete, microscopic system representation. The interaction rules in the LMA emphasize especially a correct treatment of interactions at the atomic scale.

The Lattice Molecular Automation, introduced in this paper, is a novel simulation environment suited to handle large scale molecular self-assembly problems.

sets T , F , K , P , S , and I :

In the case of molecular systems, the object complexity C is defined by the following

The basic idea behind discrete field automata is to model both, matter and forces, as mediating information particles. Three main steps determine the molecular dynamics: (i) rules that propagate force information particles, (ii) rules that evaluate the received transformation together with the local state, and (iii) rules following the update schedule to transform the system into the next time-step (moving the molecules).

2 Lattice Molecular Automation - Basic Concepts

The molecular objects themselves are encoded as data structures, defining atoms or whole molecular units (as e.g. monomers in polymers). A data structure consists of k variables, which store information as the molecular type, type specific (physico-chemical) properties defining intermolecular interactions, kinetic energies and possible formation of bonds to other molecular objects. A detailed description of a LMA data structure is given in the next section.

These interactions, hydrogen bonds, dipole-dipole - dipole (including charge-charge), dipole-induced dipole, induced dipole - induced dipole interactions and cooperativity effects, are the set of possible intermolecular potentials in the LMA. Intramolecular terms as potentials for bonds, angles (or torsional angles in 3D) are not considered. However, the contributions noted in equation (1) have proved to be essential for a correct description of macromolecular systems [1, 14]. Especially solute - solvent interactions (as e.g. polymers in aqueous environment, see also section 4.2) have to be considered to model enthalpic and entropic flows [1].

$$V_{total} = \sum_n \sum_9 \sum_{i=1}^q V_{H-bond} + \sum_n \sum_9 \sum_{i=1}^q V_{dip.-ind.dip.} + \sum_n \sum_9 \sum_{i=1}^q V_{ind.dip.-ind.dip.} + \sum_n \sum_9 \sum_{i=1}^q V_{coop.} \quad (1)$$

The contributions to a total potential energy, V_{total} , in a system with n molecules on a hexagonal lattice are obtained through:

update.

x_k . This sub-time indicates the sequence in which a particular register is having its $t_s \dots s$ is the "sub-time" step (s runs from 1-18) for updates of the individual variables \dots at actual time.

X_k , having the variables x_k on the lattice location (i,j) as arguments. $f_1(x_{k,j}) \dots$ is function 2 out of the first part of the update cycle (f_i) with the range $f_{1,2}(x_{k,j}) \dots$ denotes the variable x_k at the neighboring position $q(i,j)$ from the current lattice site (i,j) .

$x_{k,j}$... denotes the variable x_k at the lattice position (i,j) from the current sub update step at the time t .

$x_{k,j}(t_s)$... denotes slot q of variable x_k at the lattice position (i,j) , which is in the s^k element or variable in set number k .

X_k ... set number k in the data structure (object state variable space); $k = 1, 2, \dots, 7$. $q, l \dots$ principle directions on a lattice.

The variables used throughout the paper are denoted as follows:

constraints S , define the molecular dynamics. P , kinetic energies K , and even polarizabilities. Besides the potential energy situation P , kinetic energies K , and specific definition of a variety of molecular interactions, as e.g. hydrogen bonds or P , is now only based on T and F . This formulation allows, as we shall see, a highly set is together with L responsible for the construction of a type specific force field in the vicinity of the molecular entities. The calculation (using L) of potential energies, dropophobic monomers in polymers, see section 4.2) is a type specific force list F . This directly related to the molecular type T (e.g. water molecules or hydrophilic and hy-

the update cycle as we shall see in the next section.)

I ... set of interaction rules that transfer, evaluate, and transform information between and at each molecule (and vacuum) (f_i , $i = 1, \dots, 6$, since there are 6 main steps in

S ... set of constraints (excluded volume, bonds in polymers).

P ... set of accessible potential energies.

K ... set of accessible kinetic energies.

F ... set of forces between molecules (force particles).

T ... set of molecular types.

$$(2) \quad \mathcal{D}_{(i,j)} = (\mathcal{X}_{(i,j)}, \dots, \mathcal{X}_{(i,j)})$$

A data structure (object) at the lattice location (i, j) , at time t , $\mathcal{D}_{(i,j)}$ with the k variable sets $\mathcal{X}_i, i = 1, \dots, k$

Formally, these T sets can be divided into (i) subsets characterizing the molecular entities itself, \mathcal{X}_1 , i.e. the type-state \mathcal{X}_1 , the send-state \mathcal{X}_3 , the kin-state \mathcal{X}_5 , and the bound-state \mathcal{X}_7 ; (ii) into sets storing information about the states of the neighborhood, i.e. the rec-type \mathcal{X}_2 , and the rec-state \mathcal{X}_4 ; and (iii) finally a set \mathcal{X}_6 handling variables computed by \mathcal{X}_1 and \mathcal{X}_N - the move-state \mathcal{X}_6 .

$\mathcal{X}_1 = \{x_{1,1}, x_{1,2}, \dots, x_{1,q}\}; x_{1,i} \in \{0, 1\}; S; \text{bond-state}$:
 List of net energetic state (including potential and kinetic energies).
 $\mathcal{X}_2 = \{x_{2,1}, x_{2,2}, \dots, x_{2,q}\}; x_{2,i} \in \mathbb{Z}; P + K; \text{move-state}$:
 Local kinetic energy on location (i, j) in q directions.
 $\mathcal{X}_3 = \{x_{3,1}, x_{3,2}, \dots, x_{3,q}\}; x_{3,i} \in \mathbb{Z}; F; \text{send-state}$:
 Molecular types (including vacuum) in neighborhood 1 of site (i, j) .
 $\mathcal{X}_4 = \{x_{4,1}, x_{4,2}, \dots, x_{4,q}\}; x_{4,i} \in \mathbb{Z}; F; \text{rec-state}$:
 Outgoing force particles along q lattice sites.
 $\mathcal{X}_5 = \{x_{5,1}, x_{5,2}, \dots, x_{5,q}\}; x_{5,i} \in \mathbb{N}_0; K; \text{kin-state}$:
 Incoming force particles from q lattice sites.
 $\mathcal{X}_6 = \{x_{6,1}, x_{6,2}, \dots, x_{6,q}\}; x_{6,i} \in \mathbb{Z}; P + K; \text{move-state}$:
 Local kinetic energy on bonds within polymers.

insert figure 1 here

A data structure \mathcal{D} is defining all molecular (lattice) objects O through the T variables $x_1 - x_T$, which are elements in the T sets $\mathcal{X}_1 - \mathcal{X}_T$. These sets define the local objects structure (i, j) in the LMA, as also shown in figure 1:

$t + 1 \dots$ time after all sub steps in time t .

- 1.: propagation of the molecular type and redistribution of kinetic energies.
2.: construction of type-specific force fields.

Each full update cycle consists of the following main steps:

Figure 2 shows schematically an amphiphilic pentamer (four hydrophobic monomers plus one hydrophilic head monomer) on a hexagonal lattice. Different molecular surfaces of the pentamer are polar solvent molecules (like water), as well as hydrophobic monomers. The numbers indicate the types of interactions between the molecular entities, as discussed in section 3.3.

insert figure 2 here

The following considerations assume a simulation on a hexagonal lattice, i.e. with six neighboring sites, $i = 1, 2, \dots, 6$ (or corresponding to the directions $\ell = 1, 2, 3, -1, -2, -3$), but the notation holds formally for arbitrary lattice types.

3 The update cycle

The next section describes in detail the individual interaction functions (or groups of interactions, $i = 1, \dots, 6$) f_i , operating at the different sub steps ($s = 1, \dots, 18$) in the update.

The simulation takes place on a hexagonal, two-dimensional lattice, which proved to be suitable to avoid unisotropic effects when constructing discrete fields [11]. However, the general formalism holds for arbitrary lattice topologies (and is e.g. easily transformable to a cubic lattice). The simulation objects (i.e. molecular entities) and vacuum are encoded as data structures, located at each site of the lattice. Kinetic energies as well as potential energies (based on discrete force fields) are implemented to describe the molecular dynamics.

$$\mathcal{D}_{i,s}^{t+1} = \mathcal{F}(\mathcal{X}_I^{i,s}(t), \mathcal{X}_N^{i,s}(t), \mathcal{X}_O^{i,s}(t)). \quad (3)$$

is updated by an only implicitly given interaction function \mathcal{F} in the following way:

The new update cycle starts with the information propagation. This and all other updates in the sub time steps $t_1 - t_{18}$ occur in parallel from all occupied lattice sites except for t_{17} , the polymer update, which is scheduled differently (see section 3.5).

tion of kinetic energies

3.1 Step 1: Propagation of molecular types and redistribution

which means that the new value of a given variable at a given lattice site is a composed function of the variables at the site (i, j) itself and of the variables at the neighboring sites (in the six principal directions) $d = 1, 2, \dots, 6$.

$$(5) \quad f_d(x_6^b, \dots, x_1^b) = f \circ f_d(x_1^b, \dots, x_1^b)$$

The principal structure of each step in the update cycle is a function that updates one of the data structure variables x^b . The function, denoted by formal compositions "o" for the individual elements x^b on a hexagonal lattice is of the form

$$(4) \quad F = \{ f_6(x_6^b, \dots, x_1^b) \\ f_5(x_5^b, \dots, x_1^b) \\ f_4(x_4^b, \dots, x_1^b) \\ f_3(x_3^b, \dots, x_1^b) \\ f_2(x_2^b, \dots, x_1^b) \\ f_1(x_1^b, \dots, x_1^b) \}$$

In analogy to this 6-step update of the molecular system, the corresponding functions f out of the set of functions F have the following arguments:

- 6.: move the molecule and clear old lattice position.
- 5.: readjustment of bonds in polymers according to the move direction.
- 4.: calculation of the most proper move direction.
- 3.: calculation of potential energies.

$$(8) \quad \begin{aligned} & (x_{(i,j)}^{5,l-1}(t_2) + 1, x_{(i,j)}^{5,l}(t_2) - 2, x_{(i,j)}^{5,l+1}(t_2) + 1) \text{ if } x_{(i,j)}^{5,l}(t_2) < \Delta x^5 \\ & = (x_{(i,j)}^{5,l+1}(t_3), x_{(i,j)}^{5,l}(t_3), x_{(i,j)}^{5,l-1}(t_3)) = f_{1,3}(x_{(i,j)}^{5,l}(t_2), \Delta x^5) \end{aligned}$$

To consider the fact that molecules are not hard spheres, an “internal” diffusion term is implemented, spreading kinetic energies over neighboring degrees of freedom, $l+1, l-1$, if the kinetic energy in a certain direction exceeds a threshold value:

3.1.3 Internal “thermal” diffusion

This model corresponds to a classical hard sphere collision and corresponds to the r^{12} term in the Lennard-Jones potential. For simplicity, equal mass is assumed for all molecular objects.

$$(7) \quad x_{(i,j)}^{5,l}(t_2) = f_{1,2}(x_{(i,j)}^{5,l}(t_1), x_{(i,j)}^{5,l}(t_0)) = +x_{(i,j)}^{5,l}(t_0), \text{ if } x_{(i,j)}^{5,l}(t_1) \neq 0$$

A collision status, which is the basis for the exchange of kinetic energies, is indicated by values $\neq 0$ in the rec-type list (see above). In this case, the kinetic energy, stored in the *kin-state register* \mathcal{X}^5 , is transferred to these next neighbor objects.

3.1.2 Redistribution of kinetic energies

After this update, each object at location (i,j) knows whether there are any molecules in the neighborhood 1 (NHI) as well as their types. This propagation is the basis for a redistribution of kinetic and potential energies.

$$(6) \quad x_{(i,j)}^{2,-l}(t_1) = f_{1,1}(x_{(i,j)}^{1,l}(t_0)) = x_{(i,j)}^{1,l}(t_0).$$

Molecular types denote either atoms (carbon, nitrogen, ...) or whole molecular units as e.g. monomers in polymers. Each molecular object (and also vacuum) is encoded by an integer number in the first variable of the data structure from the set \mathcal{X}_1 at location (i,j) . This type-state is in a first step at t_1 propagated to neighborhood 1 by an integer number in the first variable of the data structure from the set \mathcal{X}_2 at all l directions, i.e. into the lattice positions (i,j) , $l = 1, 2, 3, -1, -2, -3$:

3.1.1 Propagation of excluded volume particles (and molecular types)

This rule ensures neighborhood 1 propagation. This propagation step is also denoted schematically in figure 2 (B). Note that propagation in direction i is stored in the directional slot $-l$ of the neighboring data structure (i,j) . This is important for the calculation of potential energies, as shown in the next section.

Propagation to $NH > 1$ is more complicated, the respective update functions are shown in the Appendix.

$$(6) \quad x_{l(i),j}^{4,-l}(t_4) = f_{z_{i,j}}(x_{i,j}^{3,l}(t_0))$$

in the respective rec-state variable in the \mathcal{X}_4 set of the data structures at the lattice site (i,j) . These site directed force particles are now propagated to the neighborhood and stored and sign, which are stored in the send-state variable from the \mathcal{X}_3 set at position (i,j) . The force field itself is decomposed into "force particles", characterized by their value hydrophilic and hydrophobic monomers or water molecules (see also section 4.1, 4.2). These force fields are capable of representing some fundamental properties of e.g. ions, of molecular entities as e.g. charges, dipoles, or the ability to form hydrogen bonds. Type specific force fields are used to characterize distinct physico-chemical properties in the respective rec-state variable excluded volumes are ensured) and kinetic energies are exchanged.

After these three operations out of the function group f_z in the sub time steps $t_1 - t_3$ the molecular neighborhood is known to each data structure on the lattice: collision states are indicated (and therefore excluded volumes are ensured) and kinetic energies are exchanged.

Kinetic energies are handled in the LMA. Collision rules in the Lattice Gas Automata [1] despite the fact that various values for the kinetic energy is similar to the (stochastic) momentum redistribution step in the as librational or rotational energy on the molecule. The result of the redistribution of This internally (i.e. on the molecular object itself) distributed energy can be interpreted $\Delta x^6 \dots$ "diffusion threshold"

sign:

This is done by repeating the propagation of \mathcal{X}^3 (see section 3.2), but with opposite molecule rotates, the particles propagated in step 2 (equation 9) have to be cancelled. orientation of the molecule (and of its chemically distinct sites) changes. Before the A possible rotation implies a new propagation step of force particles, as the relative

$$\begin{aligned} & f_{3,i}(x_{i,j}(t_0), x_{i,j}(t_4)) = \min\left[f_{3,i}(x_{i,j}(t_0) - \frac{3}{\pi}, x_{i,j}(t_4)), f_{3,i}(x_{i,j}(t_0) + \frac{3}{\pi}, x_{i,j}(t_4))\right]; \\ & \quad (11) \end{aligned}$$

The rotation operation is defined by rearrangement of dipoles in an electrostatic field. the position with lowest potential energy. This procedure corresponds formally to the It is assumed that molecules are able to rotate $\pm \frac{\pi}{3}$ and that they will be trapped in

3.3.2 Molecular rotation

2A, between two hydrophilic monomers). The strict analogy to the Coulomb term is given as $x_{i,j}^{3,i}$ and $x_{i,j}^{4,i}$ can be interpreted as two charges q_1 and q_2 normally appearing in the Coulomb term. Typical representation of this interaction are hydrogen bonds (number 3, figure 2A, between a polar solvent and a hydrophilic monomer) and dipole-dipole interactions (number 2, figure 2B, between two hydrophilic monomers).

and on incoming force information from neighboring objects at $i(i, j)$. The respective values are stored in the move-state variable \mathcal{X}^6 . The potential energy is therefore based on site directed information on the molecule on position (i, j) itself, which is the part of the Coulomb potential given by the electric charges. The function $f_{3,i}$ computes the product of the values stored in type-state, \mathcal{X}^3 , and rec-state, \mathcal{X}^4 and

$$x_{i,j}^{6,i}(t_5) = f_{3,i}(x_{i,j}^{3,i}(t_0), x_{i,j}^{4,i}(t_4)) = x_{i,j}^{3,i}(t_0) \times x_{i,j}^{4,i}(t_4), \quad (10)$$

tion step itself (see Appendix):

The calculation of potential energies follows a Coulomb term and the distance dependence and the relative dielectric permittivity ϵ_r are already considered in the propagation.

3.3.1 Interactions based on the force particle propagation

3.3 Step 3: Calculation of potential energies

$$0 = \left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{3,i} - \frac{I \nabla}{\left| \begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right|} - \left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{4,i} = \left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{3,i} - \mathbf{f}_{3,i} = \left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{4,i} - \mathbf{f}_{4,i} \quad (14)$$

one carrying the partial charge (variable $x^{4,i}$), are involved: metric, therefore both molecules, the one having the dipole (variable $x^{3,-i}$) and the wave-state variable in this particular direction. This potential is of course also symmetric, and a dipole - induced dipole interaction results. This attraction molecule surface, and a dipole - induced dipole interaction results. This is indicated by a value zero in the *send-state variable*), a partial charge is induced on the force particles stored in the *rec-state variable*) meets an uncharged molecular surface (represented by the electrons on a molecular surface. If a strong electrostatic field (represented by the interactions on a molecular surface. Inducability is an electronic property, based on the inhomogeneous local distribution of

3.3.3 Inducability

of course still local) potential energy minimum. The rotation step is only performed once for each molecular object to get a lower (but Appendix.

After this step, the function $f_{3,i}$ is executed again (time step $t_{6,m+1}$) to calculate the new potential energy situation. The general syntax for propagation to NH m is given in the steps, the distance m of the force particle propagation. After m propagation is defined by the distance m of the force particle propagation). As rotating molecules change the force field in their vicinity, also the potential energies have to be recalculated in the neighborhood of the rotating molecule (the neighborhood

$$\left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{3,i} = \mathbf{f}_{3,i} \quad (13)$$

Finally, the force particles are propagated along the new sites $l + 1$ or $l - 1$: state variables by $-\frac{3}{2}$ or $\frac{3}{2}$ to get the lowest values when summing $\sum_{i=1}^9 x^{6,i}(t_6)$ in the move-state register), transforming $X^3(t_0)$ into $X^3(t_6)$. After this step the rotation of molecules is performed (formally by adjusting the *send-*

$$\left(\begin{pmatrix} t_9 \\ t_{10} \end{pmatrix} \right)_{(t_9)}^{\top} x^{4,i} = \mathbf{f}_{4,i} \quad (12)$$

A typical example of cooperativity is the increasing strength of hydrogen bonds formed $\Delta C_{x_1} \dots$, "cooperativity volume", depending on the molecular type x_1 .

$$\frac{\Delta C_{x_1}}{(\Delta C_{x_1})_{(x_1)}} = f_{3,4}(x_{(i,j)}^{(t_0)}, x_{(i,j)}^{(t_1)}) = ((x_{(i,j)}^{(t_0)}, x_{(i,j)}^{(t_1)})^T)^{-1} x_{(i,j)}^{(t_1)} \quad (16)$$

Cooperativity effects occur when at least two values < 0 exist in the $x_{(i,j)}$ variables, i.e. at least two interactions (bonds) are formed. For each interaction direction (dynamic bond direction) the following potential energy correction is added

3.3.5 Cooperativity, n-body terms

for processes as molecular phase separation [1].

Interactions for e.g. the generation of hydrophobic effects, which in turn are a necessity between two hydrophobic monomers. These types of weak interactions proved to be between interaction number 5 in figure 2A denotes an induced dipole - induced dipole interaction $x_{(i,j)}^{(t_1)}$

$\Delta P_{x_1, x_2} \dots$ polarizability of the molecular type x_1 given a molecular neighbor of type

$$\Delta P_{x_1, x_2} = \begin{cases} 0, & \text{else,} \\ -\Delta P_{x_1, x_2}, & \text{if } x_{(i,j)}^{(t_1)} = x_{(i,j)}^{(t_2)} = 2 \end{cases} \quad (15)$$

Another type of interaction of the van der Waals type are induced dipole - induced dipole interactions, caused by uncharged surfaces in close contact (e.g. two hydrophobic molecules). The strength of this interaction is defined by the molecular type on the position (i,j) , x_1 , and on the neighboring types (i,j) , stored in $x_{(i,j)}^{(t_2)}$. The binding energy contribution is again summed on the values in the move-state variable:

3.3.4 Diffuse interactions based on polarizabilities

The contribution to ΔI_{x_1} is a fraction of the incoming force particles x_4 . The strength of the interaction is based on the inducibility of the molecular object, x_1 . Interaction 4, figure 2A, indicates a dependence on the type of the molecular object, x_1 . Dipole - induced dipole interaction between a polar solvent molecule and a hydrophobic monomer.

$\Delta I_{x_1} \dots$ inducibility of the molecular type x_1 .

excluded volumes:

In both the monomer and polymer case, one condition has to be fulfilled to ensure

$$(19) \quad d_{move}^{(i,j)}(t_{11}) = l = \max(x_{6,1}, \dots, x_{6,7}, \dots, x_{6,6}).$$

At time step t_{11} values > 0 in the move state variable in direction l indicate a force on the molecular object to move in this direction l . The largest value in the move state variable is selected as the move direction $d_{move}^{(i,j)}(t_{11})$.

$$(18) \quad \begin{aligned} & ((x_{6,i}^{(j)}(t_{10}) + x_{6,-i}^{(j)}(t_{10})) - (|x_{6,i}^{(j)}(t_3)| + |x_{6,-i}^{(j)}(t_3)|)) \\ & = f_{4,i}^{(j)}(x_{6,i}^{(j)}(t_{11}), x_{6,-i}^{(j)}(t_{10})) \end{aligned}$$

The most proper move direction is defined by the values in the *fin-state* and the move state list (storing at the time step t_{10} the final potential energies).

In a first step, the corresponding values of X_6 and X_{-6} in direction l are summed and the respective values in the opposite direction $-l$ are subtracted.

3.4 Steps 4: Calculation of most proper move-direction

Having molecular types and force particles as arguments.

At this time step all contributions to the potential energy are calculated by functions f_{10} all values $x_{6,i}^{(j)}$ are negative.

Positive values in direction l and subtracted from the value in direction $-l$. After this step at cleared on direction l and subtracted from the value in direction $-l$. After this step at two dipoles with equal sign pointing towards each other). These values are therefore caused by positive values in direction l indicate a repulsion from this direction (e.g. caused by

$$(17) \quad x_{6,i}^{(j)}(t_{10}) = f_{3,5}(x_{6,i}^{(j)}(t_9), x_{6,-i}^{(j)}(t_9)) = -x_{6,-i}^{(j)}(t_9), \text{ if } x_{6,-i}^{(j)}(t_9) < 0$$

This rearrangement takes place in the move-state variable to take care of attracting stored in the direction $-l$ are interpreted as attracting forces in direction l .

(negative) and repelling (positive) values of $x_{6,i}^{(j)}(t_9)$. Repelling forces (positive values)

3.3.6 Rearrangement of attracting and repelling forces

increase at t_9 is defined as a fraction of the bond-strength $x_{6,i}^{(j)}$ already formed at t_8 .

between water molecules, when already one H-bond is formed. In the LMA, this

$$x_{l(i_5)}(t_5) = f_{4,5}(x_{l(i_5)}(t_4), x_l(t_0)) = -\max(x_{l(i_5)}(t_4), 0) \quad (23)$$

Now, the largest value in the move-state variable of all empty lattice sites ($x_i = 0$, i.e. sites which just received $x_{(i,j)}^{6, \text{move}(t_{13})}(t_{11})$) is propagated to all sites in NHI and subtracted from the values in the respective move-state variables:

The last step in calculating the move direction is a cross-check in NH2 to avoid that more than one molecule occupies the same (new) lattice site in $t + 1$. First, the value of $x_{(i,j)}^{6,move(t+1)}$ is propagated to the (empty) lattice site $d_{(i,j)}^{move(t+1),j}$ and stored in the respective move-state variable in the slot $-d_{(i,j)}^{move}(t+1)$:

This function $f_{4,3}$ ensures, that bonds are not broken, when a monomer in a Polymer moves along the site $d_{(i,j)}^{move}(t_{12})$. If the above conditions are not fulfilled, the move direction is set = 0.

$$d_{(i,j)}^{move}(t_{12}), \text{ if } x_{(i,j)} = 1 \vee |i_{bound} - i_{d_{(i,j)}^{move}(t_{11})}| \leq 1 \vee |j_{bound} - j_{d_{(i,j)}^{move}(t_{11})}| \leq 1 \quad \text{else.} \quad (21)$$

For polymers another condition has to be fulfilled to avoid bond-breaking (the bond interaction is indicated by the number 1 in Figure 2A): Monomers in a polymer have to be always in NH1 of their connected monomers. The lattice locations (i_{bond}, j_{bond}) denote(s) the site(s) of the bonded neighbor(s) of the monomer at position (i, j), which are indicated by values = 1 in the bond-state register.

If the desired new lattice location ($i_{\text{move}}^{(t+1)}, j_{\text{move}}^{(t+1)}$) is already occupied by a molecule (see section 3.1.1). In this case the move direction is set equal to zero (i.e. the molecule (see section 3.1.1). In this case the move direction is set equal to zero (i.e. the molecule remains at the lattice location (i, j)).

$$d_{(i,j)}^{move}(t_{12}) = \begin{cases} f_{4,2}(x_{(i,j)}^{6,t}(t_{11}), x_{(i,j)}^{2,t}(t_{11}), \\ d_{move}(t_{11}), \text{ if } x_{(i,j)}^{2,d_{move}(t_{11})}(t_{11}) \\ 0 \text{ else.} \end{cases} \quad (20)$$

The parallel update of an extended object (as a polymer) on a discrete lattice only often applied routine is to update NH₂ monomers in a polymer at one time step, and along the backbone of the polymer, no completely parallel update is possible. One their next bonded neighbors. Without additional propagation of information particles is not possible [18]. Monomers in a discrete representation as in the LMA only know strict parallel update, (ii) strict local rules, and (iii) a strict conservation of momentum, based on local rules poses the following problem: The simultaneous fulfillment of (i) a strict parallel update, (ii) strict local rules, and (iii) a strict conservation of momentum is not possible [18].

The interaction functions (lookup tables) $f_{5,1}$ and $f_{5,2}$ are given in the Appendix.

$$(x_{l(i)}^{j,l}(t_{17}), x_{l(i)}^{j,l}(t_0), d_{move}^{(i,j)}(t_{16})) = f_{5,2}(x_{l(i)}^{j,l}(t_{17}), d_{move}^{(i,j)}(t_{16})) \quad (26)$$

and for monomers having two bonds:

$$(x_{l(i)}^{j,l}(t_{17}), x_{l(i)}^{j,l}(t_0), d_{move}^{(i,j)}(t_{16})) = f_{5,1}(x_{l(i)}^{j,l}(t_{17}), d_{move}^{(i,j)}(t_{16})) \quad (25)$$

the update follows:

For monomers in dimers, or end-monomers in polymers (i.e. monomers with one bond)

A move of a monomer within a polymer changes the entries in its own bond-state variable $x_{l(i)}^{j,l}(t_0)$, and in the respective lists of the bond neighbors, $x_{l(i)}^{j,l}(t_0)$.

to the move direction

3.5 Step 5: Readjustment of bonds in polymers according

A move along $d_{(i,j)}(t_{16})$ indicates, that this molecule at position (*i*, *j*) has the largest value (i.e. attractive force) to occupy the new lattice location. Following the above notation, exactly this molecule, which has the strongest force acting on it will be able to occupy the new lattice site, since it now has the value = 0 in $x_{l(i)}^{j,l}(t_{16})$. All other molecules that potentially could move to and occupy the same new lattice location, but with a smaller force acting on them, have a value < 0 in $x_{l(i)}^{j,l}(t_{16})$.

The final condition for the move direction $d_{(i,j)}(t_{16})$ is now given by

$$d_{move}^{(i,j)}(t_{16}) = 0, \text{ if } x_{l(i)}^{j,l}(t_{15}) < 0. \quad (24)$$

The first two data structures $D_1(t_0)$ and $D_3(t_0)$ show two polar solvent molecules (as water) at the time step t_0 . The molecular objects are characterized by the set \mathcal{X}_1 , i.e. the type-state \mathcal{X}_1 , the send-state \mathcal{X}_3 and the kin-state \mathcal{X}_5 . The bond-state variable $x_{i,j}$ is also present at the time step t_0 .

insert figure 3 here

position (i,j) , $(i,j+1)$, see Figure 2B).

Figure 3 shows the update of two water molecules on a hexagonal lattice in NHI (po-

4.1 A simple example of LMA molecular dynamics

4 Discussion

corresponds to the new system time $t + 1$.

At this sub time step t_{18} , one full update of a general monomer is complete and therefore

$$(x_{(i,j)}^{2,t}(t_{18}), x_{(i,j)}^{4,t}(t_{18}), x_{(i,j)}^{6,t}(t_{18})) = (0, 0, 0) \quad (28)$$

cleared on position (i,j) by a function $f_{6,t}$:

In any case, the information in rec-type \mathcal{X}_2 , in rec-state \mathcal{X}_4 , and in move-state \mathcal{X}_6 , is

$$\begin{aligned} x_{d_{\text{move}}(t_{16})(i,j)}^{2,t}(t_{18}) &= x_{d_{\text{move}}(t_{16})(i,j)}^{4,t}(t_{18}), \\ x_{d_{\text{move}}(t_{16})(i,j)}^{4,t}(t_{18}) &= x_{d_{\text{move}}(t_{16})(i,j)}^{6,t}(t_{18}), \\ x_{d_{\text{move}}(t_{16})(i,j)}^{6,t}(t_{18}) &= x_{d_{\text{move}}(t_{16})(i,j)}^{1,t}(t_0). \end{aligned} \quad (27)$$

defined by $d_{(i,j)}^{\text{move}}(t_{16})$, following a function $f_{6,t}$:

kin-state \mathcal{X}_5 , and of the bond-state \mathcal{X}_7 , are moved (not copied) to the new lattice site $H d_{(i,j)}^{\text{move}}(t_{16}) \neq 0$, the information of the type-state \mathcal{X}_1 , of the send-state \mathcal{X}_3 , of the

3.6 Step 6: Move the molecule and clear old lattice location

we refer to [5, 12] and the references cited therein.

the others in a separate, interleaved time step. For a further discussion of these issues

\mathcal{X}_1 is empty (i.e. all $x_{i,j}^{(t_1)}$ are = 0), as no bonds are formed. A bond in a particular direction l would be indicated by a value = 1 in slot l of the bond-state variable \mathcal{X}_1 . The first variable \mathcal{X}_1 stores the value 1, which is associated with a certain molecular type ((0) vacuum, (1) water, (2) hydrophobic monomers, or (3) hydrophobic monomers). The variable \mathcal{X}_2 stores the force particles. Both molecules are characterized by three distinct interaction sites (direction 1, 3, -2 for molecule (1) and 2, -1, -3 for molecule 2A). The last variable with values > 0 is the kin-state \mathcal{X}_5 , indicating a hydrogen bond, $COOH$ groups (type (2) in the LMA) can form two H-bonds (see also figure 2A). The positive values hydrogen, oxygen, positive values hydrogens): Each LMA object of type = 1 can form three hydrogen bonds, $COOH$ groups (type (2) in the LMA) can form two H-bonds (see also figure 2A). The last variable with values < 0 is the kin-state \mathcal{X}_6 , indicating a kinetic energy in direction 1 for molecule (1). Without the calculation of specific interactions (and resulting potential energies), these kinetic energies would be decisive for the molecular dynamics. In the present setup, the kinetic energies follow a Boltzmann distribution. For a detailed discussion of the relationship between kinetic and potential energy see reference [1].

The first sub-update step at t_1 , executed by $f_{1,1}$ (equation 6), propagates the excluded volume particles (fields) and the molecular types x_i to NHI. The values are stored in $\mathcal{D}_2(t_1)$. The molecular position and type in the neighborhood of the two molecules is stored in this sub-time step t_1 .

The redistribution of kinetic energies (sub-time step t_2) is now based on the just propagated information, as the values $\neq 0$, stored in the rec-type variable \mathcal{X}_2 , indicate a collision status of the two molecular objects. The function $f_{1,2}$ (equation 7) distributes the value 2, stored in $x_{i,j}^{(t_2)}$ of $\mathcal{D}_1(t_2)$, to $x_{i,j}^{(t_1)}$ of $\mathcal{D}_2(t_2)$.

The function $f_{1,3}$ (equation 8), handling an internal diffusion (sub-time step t_3), is only applied, if a kinetic energy exceeds a threshold value, Δx_5 , which is in the current time step equal to 5.

The next stage of the update is the propagation of force particles in the time step t_4 , handled by $f_{2,1}$ (equation 9). The force particles for the construction of hydrogen bonds (see end-state \mathcal{X}_3) are propagated to NHI and stored in the respective rec-state bonds (\mathcal{X}_4), handled by $f_{2,1}$ (equation 10). The force particles for the construction of hydrogen bonds (see end-state \mathcal{X}_3) are propagated to NHI and stored in the respective rec-state bonds (\mathcal{X}_4), handled by $f_{2,1}$ (equation 10).

t_{12}), as these move directions are in conflict with maintenance of excluded volumes. However, function $f_{4,2}$ (equation 20) resets d_{move} to 0 for both molecules (sub-time step $D_2(t_{11})$).

object, resulting in a move-direction $d_{move}(t_{11}) = 1$ for $D_1(t_{11})$ and $d_{move}(t_{11}) = -1$ for $D_2(t_{11})$. The largest values are now selected for each molecular variable to the move-state variable from $\mathcal{X}^6 = [4, 0, 0, 0, 0]$ in $D_1(t_{11})$ and from \mathcal{X}^6 , leading to the move-state variable from $\mathcal{X}^6 = [-2, 0, 0, 2, 0, 0]$ in $D_2(t_{11})$. In a first step, the function $f_{4,1}$ is executed (equation 18) added in the time step t_{11} . In a first step, the potential energies, the move-direction is evaluated after all these considerations for the potential energies, the move-direction is evaluated with negative sign to the opposite direction -1.

Two oxygens or two hydrogens pointing towards each other in this situation (and the concomitant loss of forming a hydrogen bond). This value +4 in direction 1 would be two oxygens or two hydrogens pointing towards each other in this situation (and the concomitant loss of forming a hydrogen bond). This value +4 in direction 1 would be variable from \mathcal{X}^6 would store the value +4 after execution of $f_{3,1}$. This would indicate [-2, 0, 0, 0, 0] in the rec-state variable from \mathcal{X}^4 of $D_1(t_5)$. In this case, the move-state variable would be: a conformation [-2, 0, 2, 0, 2, 0] in the send-state variable from \mathcal{X}^3 and \mathcal{X}^5 . A case of rearrangement equation 17) is not necessary, as all values in $x_{(i,j)}^{(i,j)}$ are < 1. A case of rearrangement equation 17) is not necessary, as all values in the move-state variable from \mathcal{X}^6 (sub-time step t_{10}), moved with negative sign to the opposite direction -1.

The rearrangement of values in the move-state variable from \mathcal{X}^6 (sub-time step t_{10}), formed. interaction (indicated by the single value -4 in $x_{(i,j)}^{(i,j)}$ for $D_1(t_5)$ and $x_{(i,j)}^{(i,j)}$ for $D_2(t_5)$) is also cooperativity (sub-time step t_9 , see equation 16) is not accounted, as only one interactions (but would be considered for two neighboring hydrophobic monomers).

Polarizabilities (sub-time step t_8 , see equation 15) are not accounted for in water-water induced dipole interaction.

With the force particle value -2, and a fraction of this value is considered as the dipole $D_1(t_5)$, indicated by the value 0 in the send-state variable, is in this case confronted rec-state variable \mathcal{X}^4 stores [-2, 0, 0, 0, 0]. The unpolar surface in direction $l = 1$ of $x_{(i,j)}^{(i,j)}$, if e.g. the send-state variable \mathcal{X}^3 of $D_1(t_5)$ stores [0, -2, 0, 2, 0, 2], and the shown, but would for instance lead to a decrease of the value -1 in directional slot 1 and shift $x_{3,l}$ to [-2, 0, 2, 0].

The function group f_3 (equations 11-13) would perform the rotation operation of $\frac{\pi}{3}$ this configuration, the rotation of a molecule would not decrease $\sum_{g=1}^6 x_{6,g}$. A rotation would take place, if e.g. $x_{3,l}$ of $D_1(t_1)$ stores the sequence [0, -2, 0, 2, 0, 2]. In this case, this configuration, the rotation of a molecule would not decrease $\sum_{g=1}^6 x_{6,g}$. A rotation yields a value -4 in slot $x_{6,1}$ of $D_1(t_5)$ and a value of -4 in slot $x_{6,4}$ of $D_2(t_5)$. In

Furthermore, thermodynamic data on self-organization processes can be evaluated, as macromolecules in aqueous environment [16, 17], are well reproduced by the LMA [1]. Such entropy driven processes and other solvation effects, intrinsic for the dynamics of drophobic polymers cluster to "minimize" their contact surface to the polar solvent. (i.e. forming intermolecular hydrogen bonds to the hydration shell), hy-solvated This example in figure 4 shows one feature of the LMA setup, the generation of the emergent Hydropobic Effect [1, 15]. Whereas hydrophobic polymers maintain well times weaker than hydrogen bonds.

monomers in NH₃). Van der Waals interactions are in the present LMA setup four times stronger than hydrogen bonds. blic monomers and induced dipole - induced dipole interactions between hydrophobic bilitized by dipole - induced dipole interactions between water molecules and hydrophobic monomers (stationary interactions) and the decamer in figure 4B of ten hydrophobic monomers (each one has the ability to form two hydrogen bonds as well as dipole - induced dipole interactions) consists of ten hydrophobic monomers (aqueous) environment. The decamer in figure 4A consists of ten hydrophobic monomers (aqueous) environment. Figure 4 shows exemplarily snapshots of simulating polymer dynamics in an polar

Figure 4 here

therefore the possibility to study molecular self-assembly processes. Applying the LMA environment allows the simulation of large molecular ensembles ($> 10^5$ molecular objects) over long time scales (up in the second range) and give

4.2 Simulation of large molecular ensembles

but \mathcal{X}_t . After this sub-time step at t_{18} the update starts again at t_0 . the data structures \mathcal{D}_1 and \mathcal{D}_2 , into the next time step $t+1$ by clearing all information would be zero, for $\mathcal{D}_2(t_{11})$ one. The functions $f_{6,1}$ and $f_{6,2}$ (equation 27,28) transform by the hydrogen bond. Considering only kinetic energies, the move direction for $\mathcal{D}_1(t_{11})$ For our situation, both molecules remain on the same lattice location. They are fixed (equations 22,23) are employed for all molecules with $d_{move} \neq 0$.

location in the next time step. To avoid such conflicts, the functions $f_{4,4}$ and $f_{4,5}$ and of the second molecule-2. Both molecules would therefore occupy the same lattice The cross check in NH2 (sub-time step t_{14}) is not performed (as both move directions are zero). This cross check would e.g. take place, if $d_{move}(t_{13})$ of molecule (1) is -3, and of the second molecule-2. Both molecules would therefore occupy the same lattice The cross check in NH2 (sub-time step t_{14}) is not performed (as both move directions are zero).

needed to construct higher order molecular systems with predefined functionalities. Simulations would specify the minimum object complexity of the molecular objects with other methods, as molecular manipulators, to gain positional control [21]. LMA take a bottom up approach for molecular design, avoiding a variety of problems occurring appear in particular fascinating: Site specific molecular self-assembly processes could besides these applications in basic research, contributions to Nanotechnology [21, 22] [21, 22] [21, 22].

ultra-ensembles.

reach a computational performance needed for large scale simulations of macromolecularism together with the more aggregated level of description are the key steps to information particle propagation steps between molecular objects. This inherent parallelism communication on the parallel machine is in complete agreement with the node or group of objects - can be directly interpreted as a computational element (node). Tomata Machine [20] or the Connection Machine: Each simulation object (i.e. molecule) implementation of the LMA on massively parallel architectures as the Cellular Automata, the concept of data structures encoding molecular objects allow a direct Furthermore, problems in the formulation of proto-life forms [1, 19, 5].

can also be used to guide experiments to clarify physico-chemical as well as conceptual minimal objects defining the solvent molecules and the monomers. Such an approach the solute and solvent molecules as water, can be generated from the low-level of the LMA. Specific interactions between macromolecules, and, most important, between highly specific functionality, as e.g. vesicles and membranes, can be treated by the especially molecular self-assembly, as found in the formation of polymer clusters with environment suitable for the treatment of large scale problems in molecular systems. Discrete field automata, as the *Lattice Molecular Automation*, provide a novel simulation

4.3 Outlook

[5].

and two successive emergent levels of a dynamical hierarchy are reported in reference formation of ordered (membrane-like) polymer clusters are reported in reference [1] in these discrete field automata systems. General phase separation processes up to the Besides such emergent properties, constructive molecular dynamics can be simulated accessible [1].

molecular states for the evaluation of entropic and enthalpic contributions are easily

$$(31) \quad x_{(i,j)}^{6,l}(t_{4,a}) = f_{2,a}(x_{(i,j)}^{4,l}(t_{4,a})) + x_{(i,j)}^{6,l}(t_{4,b})$$

Next, the new values $\neq 0$ (coming from NH2) in the rec-state variable are again added on the move-state variable:

$x_1 \dots$ reflects the dielectric constant corresponding to a molecular neighborhood x_1 .
aggregation of the force particles).

to the NH-distance from the source of the field (counted by the time steps in the prop-
er $r_m \dots$ models the distance dependence of the Coulomb potential as being proportional

$$(30) \quad \frac{r_m \epsilon_{x_1}}{x_{(i,j)}^{4,l}(t_{4,a})} = f_{2,a}(x_{(i,j)}^{1,l}(t_0), x_{(i,j)}^{4,l}(t_4))$$

constant [16] based on ‘shielding’ of the electrostatic field by the molecular type x_1 .
values in this register are propagated. This models a distance dependent dielectric
(not added) into the neighboring rec-state variable. Furthermore, only a fraction of the
Now the information in list \mathcal{X}_4 (instead of the values in \mathcal{X}_3) are propagated and written

$$(29) \quad x_{(i,j)}^{6,l}(t_{4,a}) = f_{2,a}(x_{(i,j)}^{4,l}(t_4)) = x_{(i,j)}^{4,l}(t_4)$$

move-state variable.

First, the information in rec-state, \mathcal{X}_4 , is copied into the, up to this time step empty,
neighbors involves additional steps:

Propagation to NH1 in time step t_4 is discussed in section 3.2. Propagation to the next

5.1 Propagation of force particles to neighborhood < 1

5 Appendix

B. M. thanks the *Fonds zur Förderung der wissenschaftlichen Forschung in Austria* for generous financial support (project number P09750-CHE) as well as the Los Alamos National Laboratory and the Santa Fe Institute for their hospitality.

Acknowledgements:

from 1 to -3 and concomitant on the neighboring monomer from -1 to 3. If the direction 2 is the most suitable, the bond information on location (i,j) is changed configuration, a move of the monomer on location (i,j) in direction 2 or -3 is allowed. bond monomer on location (i,j) must have a value = 1 in direction -1. In this Example: If the monomer on location (i,j) has a bond in direction 1, the neighboring updated from the sub-time step t_0 to t_7 . Following the rule table, the bond information on location (i,j) and location (i,j) is possible on a hexagonal lattice to ensure excluded volume and to avoid bond breaking. For monomers in dimers or end-monomers in polymers, only two move directions are (i,j) . These two entries have to be updated according to the move direction $d_{move}(t_6)$. in direction -1 of variable $x_{(i,j)}^{(t_7)}$ on the neighboring, bond monomer on the location which has to be updated. Corresponding to this direction 1 is a bond information (i.e. an entry 1 in the bond-state variable in direction 1) on the monomer information (i,j) denotes the bond This table considers monomers having one bond formed. $x_{(i,j)}^{(t_7)}$

⊖ Rule table $f_{5,2}$:

5.2 Rule tables for the readjustment of bonds in polymers

as discussed in section 3.3.
After this step, the update cycle continues with the calculation of potential energies,

$$x_{(i,j)}^{(t_4,m)} = f_{2,c}(x_{(i,j)}^{(t_6,t)}(t_{m-1})) \quad (32)$$

As a last step, all values in the move-state variable are moved up to the rec-state variable.

The move-state variable stores the received force particles from all neighborhoods, the rec-state variable stores the actual "wave front" of the force particle propagation. This cycle (sub-time steps t_4,b and t_4,c) is now repeated $m - 1$ times, where m is the desired NH distance for the force particle propagation.

$$(34) \quad x_{l(i^*)}^{t_1}(t_0) \quad x_{l(i^*)}^{t_1}(t_0) \quad d_{l(i^*)}^{move}(t_{16}) \quad x_{l(i^*)}^{t_1}(t_{17}) \quad x_{l(i^*)}^{t_1}(t_{17})$$

(3; 1) (-3; -1) (2; -1) (-2; 1)
 (2; -1) (-2; 1) (3) (-3; 2)
 (3; -3) (1; 3) (-2) (-3; 1)
 (-1; -3) (1; -3) (-2) (3; 1)
 (-3; -2) (3; 2) (-1) (2; -3)
 (2; -1) (-2; 1) (3) (1; -2)
 (3; 1) (-3; -1) (2) (-1; -3)
 (2; -3) (-2; 3) (1) (3; -2)
 (-3; 2) (3; -2) (-3; 2) (2; -3)

This rule table is applied to monomers having two bonds. In this case, only one move direction is allowed for each particular bond configuration.

$$(33) \quad x_{l(i^*)}^{t_1}(t_0) \quad x_{l(i^*)}^{t_1}(t_0) \quad d_{l(i^*)}^{move}(t_{16}) \quad x_{l(i^*)}^{t_1}(t_{17}) \quad x_{l(i^*)}^{t_1}(t_{17})$$

(3) (-3) (2; 1) (-2; 1) (1; -2)
 (2) (-2) (1; -3) (-1; -3) (3; 1)
 (-1) (1) (3; -2) (-2; 3) (2; -3)
 (3) (-3) (2; -1) (-1; 2) (1; -2)
 (2) (-2) (3; 1) (1; 3) (-1; -3)
 (1) (-1) (2; -3) (-3; 2) (3; -2)

- [1] B. Meyer, G. Kohler and S. Rasmussen, *Simulation and Dynamics of Entropy-Driven, Molecular Self-Assembly Processes*, submitted to Phys. Rev. E (1996), Los Alamos National Laboratory and Santa Fe Institute preprint available from authors.
- [2] R. B. Genni, *Biomembranes*, (Springer, Berlin, 1988).
- [3] V. S. Kulkarni, W. H. Anderson and R. E. Brown, *Bioophys. J.* **69** 1976 (1995).
- [4] N. A. Baaas, in *Artificial Life III, SFI Studies in the Science of Complexity*, edited by Ch. G. Langton, Proc. Vol. XVII (Addison Wesley, 1993), p. 515.
- [5] N. A. Baaas, M. W. Olesen and S. Rasmussen, *Generation of Higher Order Emergent Structures*, Los Alamos National Laboratory and Santa Fe Institute preprint (1996), available from authors.
- [6] S. Rasmussen and Ch. L. Barrett, in *Advances in Artificial Life*, Lecture Notes in Artificial Intelligence 929, edited by F. Moran et al. (Springer, Berlin, 1995),
- [7] S. Bonacchio, M. Wessicken, D. Berti, P. Waller and P. L. Luisi, *Relation between the Molecular Structure of Phosphatidyl Nucleosides and the Morphology of their Supramolecular Aggregates*, Preprint (1996), may be obtained from the authors at the Institute for Polymers, ETH-Zürich, Universitystrasse 6, CH-8092 Zürich, Switzerland, private communication, December 1995.
- [8] P. Waller, Institute for Polymers, ETH-Zürich, Universitätstrasse 6, CH-8092 Zürich, Switzerland, private communication, December 1995.
- [9] W. F. van Gunsteren, J. F. Lange and P. Tijms, *Ann. Rev. Biophys. Biomol. Struct.*, **23**, 847 (1994).
- [10] G. Bohm, *Chaos, Solitons and Fractals*, **1**, 4, 375 (1991).
- [11] U. Frisch, B. Hasslacher and Y. Pomeau, *Phys. Rev. Lett.*, **56**, 1722 (1986).
- [12] S. Rasmussen and J. R. Smith, *Ber. Bunsenges. Phys. Chem.*, **98**, 1185 (1994).

References

- [22] R. C. Merkle, *Nanotechnology*, 2, 134 (1991).
- [21] K. E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, (John Wiley & Sons, 1992).
- [20] T. Thofoh and N. Margolus, *Cellular Automata Machines - A New Environment for Modeling*, (Cambridge, MIT Press, 1987).
- [19] T. Oberholzer, R. Wick, P. L. Luisi and Ch. Biebricher, *Biochem. Biophys. Res.* Comm., 207, 1, 250 (1995).
- [18] Y. Bar-Yam, Boston University, private communication (1995).
- [17] P. Privalov and S. J. Gill, *Adv. Prot. Chem.*, 39, 191 (1988).
- [16] C. Th. Klein, B. Meyer, G. Kohler and P. Wolschann, *Influence of Solvation on α -helix formation of Poly-alanine studied by multiple annealing simulations*, THEOCHEM, accepted for publication (April 1996), available from Bernd Mayer, Institute for Theoretical Chemistry, UZAII, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria.
- [15] Ch. Tanford, *Science*, 200, 1012 (1978).
- [14] S. K. Burley and G. A. Petsko, *Adv. Prot. Chem.*, 39, 125 (1988).
- [13] W. C. Jorgenseen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, 106, 6638 (1984).

Figure Captions

Figure 1:

Two data structures, D_1 and D_2 , characterizing molecular objects in the LMA. The seven variables $x_1 - x_7$ store type-specific information as the *type-state*, the *send-state*, the *kin-state* and the *bind-state*; information received from the neighborhood in propagation steps P , the *rec-type* and the *rec-state*; and finally information computed internally, the *move-state*.

Figure 2:

(A) Schematic drawing of LMA molecular objects on a 2D, hexagonal lattice. The polymer (pentamer) consists of four hydrophobic monomers (as e.g. CH_n groups), drawn as unfilled circles, and of a hydrophilic head monomer (as e.g. a $COOH$ group), drawn as an unfilled square. A polar solvent (as e.g. water) is represented as black circles. The numbers (1-5) denote the different types of interactions implemented in the LMA: (1) bonds between monomers in polymers; (2) dipole - dipole interaction; (3) hydrogen bond; (4) dipole - induced dipole interaction; (5) induced dipole - induced dipole interaction.

(B) Force particle propagation of two molecular objects (LMA representation of water) in NH₁ (position (i, j) , $(i, j + 1)$). Bottom: Notation of lattice directions on hexagonal lattice.

Figure 3:

Update sequence of two polar solvent molecules in NH₁ on the hexagonal lattice (as shown in figure 2B): Denoted are the data structures D_1 and D_2 , with seven variables $x_1 - x_7$, characterizing molecular objects in the LMA setup.

Figure 4:

Snapshots of LMA simulations of a hydrophilic decamer (A) and a hydrophobic decamer (B) (black filled circles) in a polar solvent (unfilled circles) after 10^4 update steps. 50 % of the lattice sites are occupied with molecular objects.

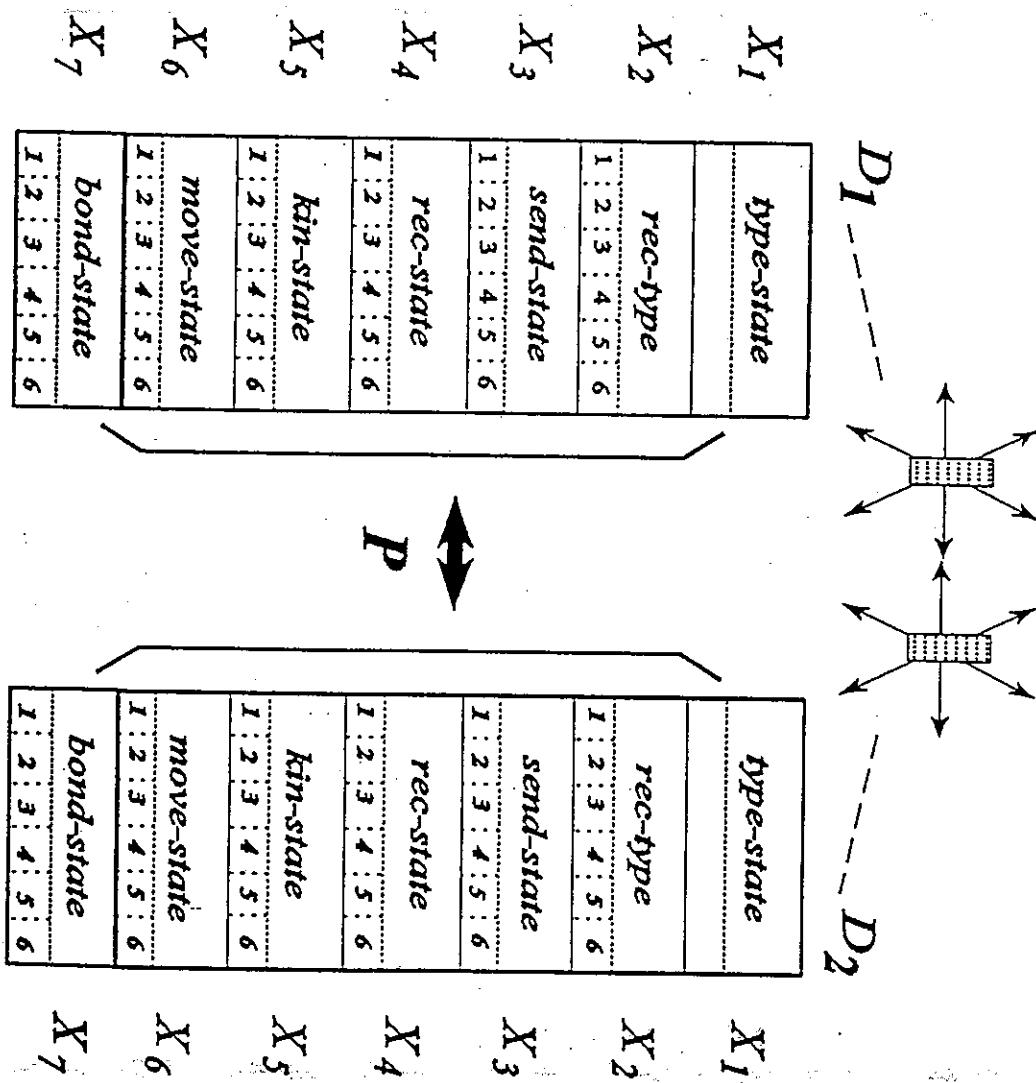


Fig. 1

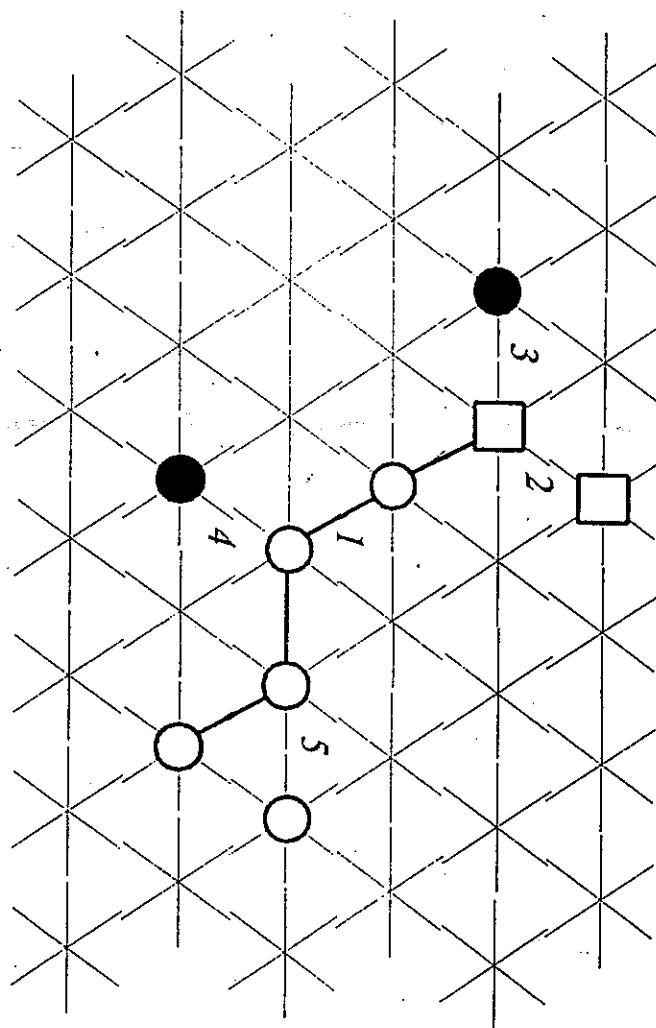
● □

hydrophilic monomer

hydrophobic monomer

polar solvent

A



● •

force particle

B

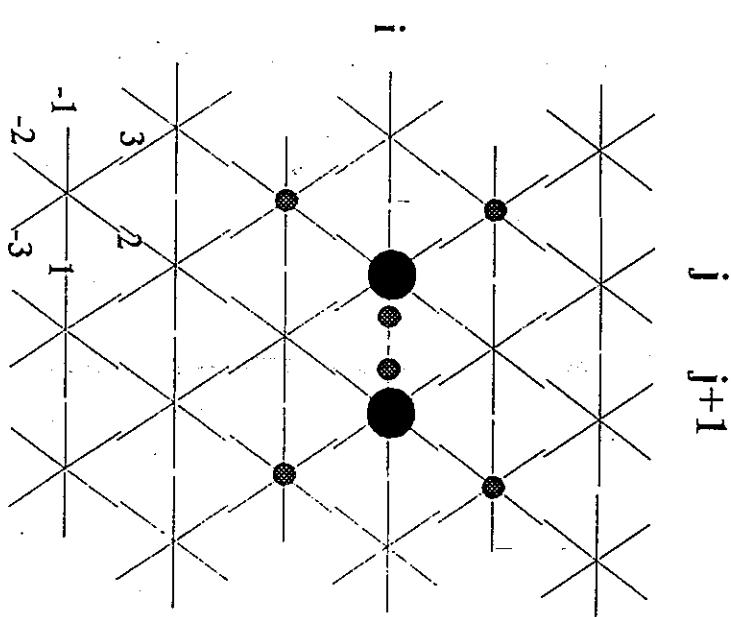


Fig. d

19.3

0:0:0:0:0:0
-2:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:-2:0:0
2:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
4:0:0:0:0:0
0:0:0:0:0:0
2:0:0:0:0:0
2:0:0:0:0:0
-2:0:2:0:0:0
0:-2:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
-4:0:0:0:0:0
0:0:0:0:0:0
2:0:0:0:0:0
2:0:0:0:0:0
-2:0:2:0:0:0
0:-2:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
-4:0:0:0:0:0
0:0:0:0:0:0
2:0:0:0:0:0
2:0:0:0:0:0
-2:0:2:0:0:0
0:-2:0:1:0:0
1:0:0:0:0:0

$\mathcal{G}_1 t_5$ $f_3,1$ $f_4,1$ $\mathcal{G}_2 t_{11}$

$\mathcal{G}_1 t_8$ $f_6,1$ $f_6,2$ $\mathcal{G}_2 t_{18}$

0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:0:0:0:0:0
1

0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:0:0:0:0:0
1

0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:-2:0:0
2:0:0:0:0:0
-2:0:2:0:2:0
0:0:0:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

$\mathcal{G}_1 t_2$ $f_1,2$ $f_2,1$ $\mathcal{G}_2 t_4$

$\mathcal{G}_1 t_8$ $f_6,1$ $f_6,2$ $\mathcal{G}_2 t_{18}$

0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:0:0:0:1:0:0
1:0:0:0:0:0

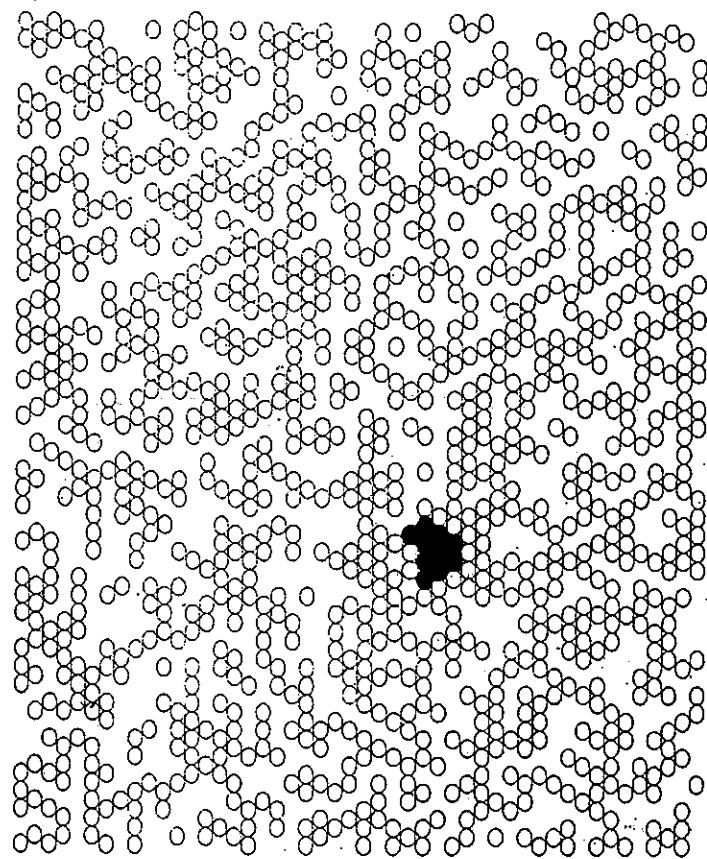
0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:0:0:0:1:0:0
1:0:0:0:0:0

0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

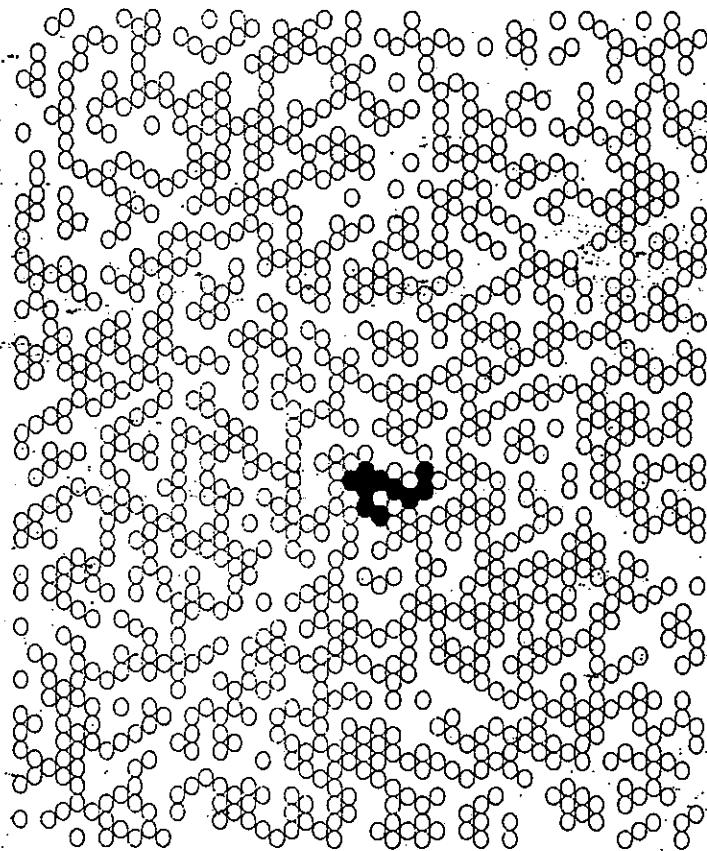
0:0:0:0:0:0
0:0:0:0:0:0
2:0:0:0:0:0
0:0:0:0:0:0
0:0:0:0:0:0
-2:0:2:0:2:0
0:-2:0:1:0:0
1:0:0:0:0:0

$\mathcal{G}_1 t_0$ $f_1,1$ $f_2,1$ $\mathcal{G}_2 t_1$

$\mathcal{G}_1 t_8$ $f_6,1$ $f_6,2$ $\mathcal{G}_2 t_{18}$



B



A

Fig. 4