ALGORITHMIZATION AND SIMULATION OF THE CHAIN-LIKE STRUCTURES' DYNAMICS-INTERRELATIONS BETWEEN MOVEMENT CHARACTERISTICS

Andrzej Z. GRZYBOWSKI, Zbigniew DOMAŃSKI, Kamila BARTŁOMIEJCZYK
Faculty of Mechanical Engineering and Computer Science, Czestochowa University of Technology, Czestochowa, Poland,
e-mail: andrzej.grzybowski@im.pcz.pl, zbigniew.domanski@im.pcz.pl, kamila.bartlomiejczyk@im.pcz.pl,

ABSTRACT
We describe new approach for algorithmization of the chain-like structures' dynamics. The main underlying idea of our approach is the sequentialization of the moves. The resulting algorithm enables efficient sampling of vast state spaces related to considered phenomena. In our simulation experiments the algorithm appeared to be much faster than other algorithms known from literature. Therefore it enables researchers to study complex models of such systems as biomolecules or artificial polymers. In this paper we perform simulation study of interrelations between various parameters specifying different movement characteristics. Also the impact of these parameters on the time taken to cover a given distance along a given trajectory in the motion space is analyzed.

1. INTRODUCTION

Chain-like structures (CLS) are common in Nature, and thus they are widely encountered in biological, physical and technological processes. An important example of such processes involves a polymer transfer through a cell's membrane which is vital to virtually all living organisms [7]. From the physical point of view polymers are complex systems that can be studied by Monte Carlo simulation [11,9]. However, the complexity of possible polymer's behaviour along with the variety of the environment space futures form a real challenge for any algorithm used for this kind of studies.

The principal aim of this presentation is to depict an innovative concept for the algorithmization of the CLS movement. The main underlying idea is to sequentialize each move of the CLS. This concept enables us to model – among others – the propagation of the tension through the whole CLS structure. Our approach significantly reduces the search space consisting of the allowed (acceptable by Nature) states. Thus, in our opinion it is obvious that the resulting algorithm is faster than those that have been presented so far in literature [2, 6,10,14]. Due to increased efficiency, the new algorithm also enables the researcher to take into account various additional features of the CLS. In this paper we use the proposed algorithm in simulation studies of the interactions between a number of algorithm parameters which are used to model numerous movement characteristics of the CLS. These characteristics are primarily related to the elastic properties of the chain.

Below, in Section II, we define basic notions and state principal assumptions which are necessary for description of our algorithm. The algorithm itself is presented in Section III. Next, in Section IV, we describe a simulation framework that we use to study some important aspects of modeling of the CLSs' dynamics. We focus on the relation between certain algorithm parameters and the speed of the CLS movement. Based on simulation experiment results we build metamodels for this realtionship. They are presented in Section V.

2. BASIC DEFINITIONS AND ASSUMPTIONS

In the formal description of the algorithm we use specific terminology which is necessary for precise denotations, which we define below.

An abstract CLS *position* is a finite sequence $\mathbf{c} = \{c_1, .. c_n\}$ of *2D* points $c_i = \{x_i, y_i\}$ such that the distance d between two successive elements is bounded by given limits D_{min}, D_{max}, i.e. $D_{min} \le d(c_i, c_{i+1}) \le D_{max}$, $i = 1, 2, \ldots, n-1$. The elements of the sequence $\mathbf{c}$ are called *segments* of the CLS.

The segment c_n is called a head of the chain, while the element c_i is called a tail. The number n is called a length of the chain.

Assumption 1 (*discretization of the motion space*)
The chain moves along the integer lattice nodes, i.e. the coordinates $\{x, y\}$ of the chain segments are integers.

It is a very common assumption in the literature, [6,10,13]. However, it is obvious that in this case to achieve better approximation of the continuous motion space, it can be assumed that the abstract unit of the length is equivalent to a given number of grid sides. The greater is the number of grid sides per unit (GSPU), the better is the approximation.

Any pair of the consecutive segments is called a bond. The structure of the CLS is defined by the mutual relations between all segments and bonds.

In the description of the algorithm we will use the notions of *steps* and *moves*. The first one is connected with a segment, whilst the second one is related to the whole CLS.

A *step* carries a given single segment from one lattice node to another. A single step can randomly transform a given segment into one from a set of its one step reachable nodes (OSRN), i.e. nodes satisfying the condition: $d(c_i, c_i') \le L_{max}$, where L_{max} defines assumed maximum length of a step made by any segment. In real-world interpretation the parameter L_{max} may reflect, e.g. the spring nature of a chain. By making a step the segment may create a tension in the CLS structure. Particularly, it may happen when $d(c_{i-1}, c_i') > Dmax$ and/or

$d(c_{i+1}, c'_i) > Dmax$. A move of the CLS consists of steps made by consecutive segments as long as the tension in the structure exists. So, the move reflects the tension propagation. If the first step does not create any tension in the CLS structure, then the move may consist of a single step only.

Assumption 2. (*sequentialization of the CLS movement*): Every move of the CLS is initialized by only one segment. Then the tension propagation trough the CLS can be sequentialized into a sequence of steps.

The first to go segment (FTG) is the segment which is chosen by the algorithm to initialize every next move of the CLS. The choice of FTG is made randomly according a given probability distribution defined on the CLS. The distribution will be denoted as FTGCD. The FTGCD may model various physical aspects of the CLS. An example of a problem, where the FTGCD is not uniform are polyelectrolytes, or, more generally, charged chains where the Coulomb interaction as well as the influence of charge fluctuations lead to models with different choice probabilities reflecting the charge volume.

A movement trajectory is a sequence of consecutive CLS positions stored in matrix C whose *i*-th row is interpreted as a CLS position after *i*-1 moves (at moment *i*). Thus symbol c_{ij} denotes the position of segment *j* in the CLS at the moment *i*.

The final CLS position after the move is made may be influenced by various physical laws. In our algorithm these lows can be incorporated into simulation model by a probability distribution defined on the OSRN set. This distribution defines probabilities of different directions and/or lengths of each step. This step probability distribution may also depend on the segment's coordinates. It will be denoted by SPD.

Some features of the environment as well as assumed properties of the CLS itself, may result in existence of actually forbidden nodes (AFN). For any given moment and for each segment AFN there exists a subset of OSRN, and, obviously, it may vary in time. For example, one of such restrictions may be a requirement that in a given node, at most a given number of segments can be placed (e.g. the repton model [13], self-avoidance restriction etc). Yet another example is the existence of other objects that occupy some nodes, e.g. a cellular membrane.

The subtraction of SRN and AFN will be called a set of actually accessible nodes (AAN): AAN=SRN - AFN. The distribution SPD truncated to the set AAN will be called actual step probability distribution and denoted as ASPD.

In many real-world physical problems, such as polymer behaviour inside a tissue of living organism, we should also take into account additional constraints related to the biochemical nature of the system under consideration. Thus we define additionally the cost connected with the polymer structure. The cost of the polymer structure and its location in the motion space is a function F representing its fitness connected with its conformation and/or other external (e.g. environmental) properties. The lower cost, the better fitness of the polymer structure and position.

Assumption 3 (CLS *position acceptance*): The new position of the CLS is accepted with a probability depending on its cost.

3. THE ALGORITHM

The above assumptions and ideas are implemented in the algorithm for CLS movement simulation. Its block scheme is presented in Fig. 1.
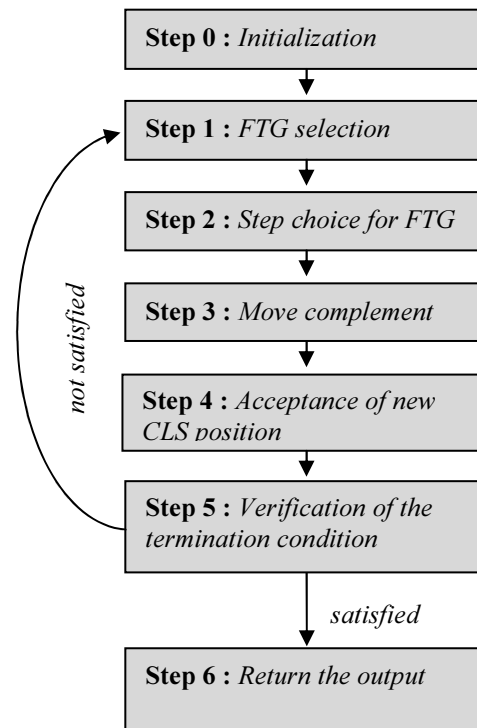

Fig. 1 Block scheme of the algorithm for CLS movement simulation

In Step 0 (initialization) the space and movement parameters (D_{min}, D_{max}, L_{max}, GSPU) are set. At the same step the user also sets the initial *chain position* c_{curr} and program evaluates its current cost function value F_C. Dependently on the simulation setup the first position of the CLS may also be chosen at random. Next, in Step 1, according to given FTGD the FTG is selected. The chosen position is denoted $c_{curr,f}$. The essence of our algorithm is the Step 3: move compliment. During this step *successive steps of remaining segments* are made to obtain a new chain position c_{new}. The algorithm chooses sequentially segments $c_{new,i}$, *i*=*f*-1,..,1 and draws randomly neighboring nods for their next position according the OLPD and AFR. This process is terminated for the first *k*, *k*=*f*-1,..,1 for which the following condition holds: $d(c_{curr,k}, c_{new,k+1}) < g$. If *k*>1, then for *i*=1,..,*k* we assume $c_{new,i}=c_{curr,i}$. Next the algorithm chooses sequentially segments c_i , *i*=*f*+1,..,*n* and draw at random their next position according the OLPD and AFR. This process is terminated for the first *k*, *k*=*f*+1,..,*n* for which the following condition holds: $d(c_{curr,k}, c_{new,k-1}) < g$. If *k*<*n*, then for *i*=*k*,..,*n* we assume $c_{new,i}=c_{curr,i}$.

In difference to all algorithms reported in literature previously, the proposed method of CLS position generation assures that it - the new position - satisfies the most fundamental chain properties. Consequently, the cost

function that is usually adopted to verify the position does not reject any one generated in our approach. However in many problems, there are also some additional properties and/or requirements that should be satisfied by the CLS position and/or structure. Thus, whether or not the new position can be finally accepted is verified in Step 4. In this step the cost F_N of the new CLS position and the difference $\delta = F_N - F_C$ are computed. If $\delta < 0$ then $\mathbf{c}_{new}$ is accepted. Alternatively, the new position $\mathbf{c}_{new}$ is accepted only if a random variable U having the uniform probability distribution on the interval [0,1] satisfies $U \le A(\delta)$. If $\mathbf{c}_{new}$ is accepted then $\mathbf{c}_{curr}$ is replaced by $\mathbf{c}_{new}$, otherwise $\mathbf{c}_{curr}$ remains as is. After generation of the new CLS position the termination condition is verified. In a given simulation experiment this condition is closely related with the investigated problem, but usually it involves some requirements that must be satisfied by the CLS position, e.g. whether the CLS has reached a given set of states or whether it has passed through a pore in a membrane. If the termination condition is met the simulation is terminated, otherwise the algorithm returns to Step 1. After the simulations are terminated, the algorithm returns the output, which dependently on the simulation purpose, may consist of the movement trajectory and its various statistical characteristics.

4. SIMULATION STUDIES AND THE RESULTS

In this section we present a simulation framework that we use for the analysis of CLS dynamics. It is assumed that the CLS is going upwards towards a boundary (which may reflect e.g. presence of cell membrane). We want to investigate the influence of the introduced above algorithm parameters on the number of steps which the CLS needs to made in order to reach the boundary - let us denote this quantity as NSB. It is a problem similar to the one investigated in [1] with the help of other algorithm, where the probabilistic distribution of the NSB where investigated. It is assumed that the probabilities that the step is made upwards are twice as big as the probabilities of a move in the opposite direction. In each case, at the initial position the CLS is placed parallel to the boundary and below it in a distance equal to 30 units. The initial position of CLS along with an example of its simulated final position is presented in Fig. 2.
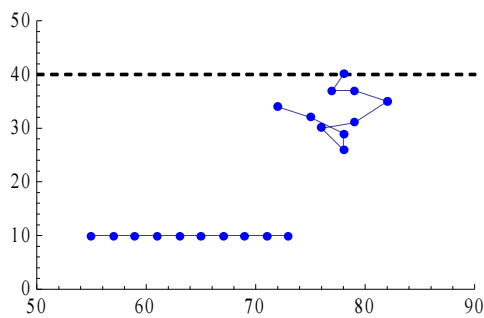

Fig. 2 An initial position of the CLS along with an example of its simulated final position - CLS length equals 10. The boundary is presented as a dashed line.

The movement parameters which reflect the CLS elastic properties are primarily the D_{max} and L_{max}. In our introductory studies we examine the influence of these parameters on the movement "time" measured in terms of the NSB. We assume the following values for other parameters used in the simulations: GSPU=2, D_{min} equals 1 unit (2 grid sides). In our simulation framework the parameter D_{max} takes on the values 2,4 or 6 units, while the parameter L_{max} takes on the values 3,6,or 9 units. The NSB obviously depends also on the CLS length. In our simulation we study three different lengths of CLS: 10, 20 and 40. Results of our simulation experiment are presented in Tables 1,2 and 3.

Table 1 Mean values of NSB in relation to the parameters D_{max} and L_{max} - simulation results obtained for CLS Length equal to 10

	$D_{max} = 2$	$D_{max} = 4$	$D_{max} = 6$
$L_{max} = 3$	291.8	413.5	437.9
$L_{max} = 6$	153.2	131.0	142.4
$L_{max} = 9$	111.7	90.2	75.9

Table 2 Mean values of NSB in relation to the parameters D_{max} and L_{max} - simulation results obtained for CLS Length N= 20

	$D_{max} = 2$	$D_{max} = 4$	$D_{max} = 6$
$L_{max} = 3$	536.1	757.9	775.5
$L_{max} = 6$	277.9	181.9	226.6
$L_{max} = 9$	161.0	122.7	93.3

Table 3 Mean values of NSB in relation to the parameters D_{max} and L_{max} - simulation results obtained for CLS Length N=40

	$D_{max} = 2$	$D_{max} = 4$	$D_{max} = 6$
$L_{max} = 3$	872.7	1302.3	1374.3
$L_{max} = 6$	404.8	269.6	366.7
$L_{max} = 9$	326.5	146.3	135.6

Based on the presented results the obvious conclusion is that the longer the CLS, the more "time" it needs to cover the distance to the boundary - for a greater length the NSB is greater in each case of the analogous values of the two parameters under study. However, studying the results presented in the above tables one may also notice some more interesting features of the examined relationship. The most important one is that the impacts of the two parameters (D_{max} and L_{max}) mutually interact. We see that the greater is L_{max}, the faster is the movement. However, the acceleration is stronger when the parameter D_{max} is less than L_{max}. But it is the necessary condition for the "spring effect". Consequently, the results prove that the impact of the spring effect is well reflected by our algorithm.

To describe this important relations in more general form we build proper metamodels.

5. METAMODELS FOR THE RELATION BETWEEN NSB, D_{MAX} AND L_{MAX}

In a number of real-world problems, as well as in various stochastic simulation analyses, the data gathered through the observations process posses the black-box structure. In the analysis of the input-output relation under the black-box representation, a special role is played by so-called metamodels [8]. A very general definition of the metamodel is the following: a *metamodel f* is an approximation of the input/output relating function ϕ that is defined by the underlying *simulation model*. Metamodels are built on the base of the simulation data with the help of various statistical techniques, first of all the regression analysis. In this section we present metamodels relating the algorithm parameters N, D_{max}, L_{max} and the movement speed characteristic *NSB*. The assumed form of the metamodel is the following:

$$NSB = f(L_{max}, D_{max}, N) + Z \qquad (1)$$

where Z is a random variable (disturbance) with E(Z)=0 and finite variance.

In our simulation experiment the parameters N, D_{max}, L_{max} where chosen from the following ranges of integers: $N \in [10,100]$, $D_{max} \in [2,20]$, $L_{max} \in [2,20]$.

It appears that the relation given by f in (1) has nonlinear character. What is especially interesting the parameters N, D_{max}, and L_{max} interact – the impact of one of these parameters depends on the values of the others. Based on the preliminary data analysis, we consider the following shape of the regression function f:

$$f(L_{max}, D_{max}, N) = \beta_0 + \beta_1 D_{max} + \beta_2 L_{max} +$$

$$\beta_3 D_{max} L_{max} + \beta_4 L_{max}^2 + \beta_5 N + \beta_6 \frac{N}{L_{max}} + \beta_7 N^2 \qquad (2)$$

The estimates b_i of the regression coefficients β_i, $i=0,...,7$, are as follows:

$b_0 = -3.83752$, $b_1 = 24.7051$, $b_2 = -33.9772$,
$b_3 = -1.59403$, $b_4 = 2.36439$, $b_5 = -12.159$,
$b_6 = 281.96$, $b_7 = -0.0496536$

It turns out that the model has really good statistical characteristics. For example its coefficient of determination equals $R^2 = 0.9800$ and *all* indicted explanatory variables have significance below $p=10^{-17}$ (usually $p=0.05$ is considered as good enough!).

The metamodel (2) describes well the character of the interrelations between the algorithm parameters and their mutual influence on the simulated CLS movement speed. It allows a proper "calibration" of the algorithm parameters in order to improve the quality and adequacy of the final model in its given real-world application.

It is also very important that the simulation results confirm our intuitions. Such confirmation of the intuitively expected features of the investigated relationship supports the adequacy of the presented approach for modeling CLS movement. Although in physics literature, particularly in polymer studies, the problem of the elasticity of the bonds

is treated as very important one, such experiments have not been reported in the stochastic simulation literature as yet. The presented algorithm enables us to study various aspects of the CLS movement when the whole structure demonstrates spring properties and the tension propagation should be taken into account.

6. FINAL REMARKS

Based on the simulation experiment described in this paper we see that the concept of *s*equentialization of the CLS movement leads to fast and effective algorithms that allow to model CLS dynamics in an efficient way. Although the CLS model investigated here was intentionally chosen in its simple version (e.g. a cost of current CLS position was not taken into account) the results received in our simulations are very important. They prove that even purely geometric algorithm parameters as D_{max}, L_{max} can be used to model some important physical features of the CLS as e.g. its elastic properties. The mutual interrelation between the model parameters and their impact on the movement speed can be studied with the help of the matamodel presented in Section V. Such studies allow the experimenter better modeling of a real-world phenomena involving numerous variation of chain-like-bodies transportation processes.

The algorithm can also be easily applied to more sophisticated physical and biological problems. Eminent examples come from the so-called DNA origami technique which allows engineers to precisely locate the ends of DNA strand on a substrate and in consequence to use the DNA as a piece of scaffolding to produce few-nanometer-resolution pattern of computer chips [5,12].

Specifically, in the spirit of the presented algorithm, it is worth mentioning that one can easily take into account the interactions among the segments (e.g. the so-called spring cost or other possible costs connected with the CLS structure alone which are important in a number of specific, biologically-oriented problems). Also the possible interactions between the segments (e,g. monomers) and the environment determining the movement space (e.g solvent's molecules) can be implemented into the simulation experiment with the help of presented here approach. Some applications of the presented algorithm to other realistic polymer's models were presented in [4]. Another polymer studies as well as some further modifications of the algorithm itself are in the progress.

REFERENCES

[1] BARTŁOMIEJCZYK, K.: "Distribution of position of segment initiating the chain-like structure translocation through the membrane opening", Journal of Applied Mathematics and Computational Mechanics, 12(1), pp. 5-10, 2013.

[2] BINDER, K.: Monte Carlo and Molecular Dynamics Simulations in Polymer Science, Oxford University Press, Oxford, 1995.

[3] DOMAŃSKI, Z. – GRZYBOWSKI, A. Z.: "An approach to polymer translocation via a sequential algorithm", Proceedings of the World Congress on Engineering and Computer Science 2012, vol. II,

WCECS 2012, San Francisco, USA, Oct. 24-26, 2012.

[4] DOMAŃSKI, Z. – GRZYBOWSKI, A. Z.: "Simulation study of translocation molecules through artificial nanopores", Proceedings of International Scientific Conference on Informatics, Informatics' 2013.

[5] DOUGLAS, S. M. – MARBLESTONE, A. H. – TEERAPITTAYANON, S. – VAZQUEZ, A. – CHURCH, G. M. – SHIH, W. M.: "Rapid prototyping of 3D DNA-origami shapes with caDNAno", Nucelic Acids Reaearch 2009, vol. 37, no. 15, pp. 5001-5006.

[6] DRZEWIŃSKI, A. – van LEEUWEN, J. M.: "Crossover from reptation to Rouse dynamics in the extended Rubinstein-Duke mode"l, Phys. Rev. E77, 031802, 2007,. doi: 10.1103/PhysRevE.77.031802.

[7] GERSHOW, M. – GOLOVCHENKO, J. A.: "Recapturing and trapping single molecules with a solid-state nanopore", Nature Nanotechnology 2, 775-779, 2007,. doi:10.1038/nnano.2007.381.

[8] KLEIJNEN, J. P. C.: Design and Analysis of Simulation Experiments, by Springer Science+Business Media, New York 2008.

[9] LANDAU, D. P. – BINDER, K.: A Guide to Monte Carlo Simulations in Statistical Physics, Cambridge University Press, Cambridge, 2005.

[10] van LEEUWEN, J. M. – DRZWIŃSKI, A.: "Stochastic lattice models for the dynamics of linear polymers", Physics Reports 475, pp. 53-90, 2009, doi: 10.1016/j.physrep.2009.04.001.

[11] SPALL, J. C.: Introduction To Stochastic Search And Optimization; Estimation, Simulation, And Control, A John Wiley & Sons. Inc., Publication, 2003.

[12] ROTHEMUND, P. W. K.: "Folding DNA to create nanoscale shapes andpatterns", Nature 2006, vol. 440, pp. 297-302.

[13] RUBINSTEIN, M.: "Discretized model of entangled-polymer dynamics", Phys. Rev. Letters, vol. 59, pp. 1946-1949, Oct. 1987.

[14] ŻUREK, S. – KOŚMIDER, M. – DRZEWIŃSKI, A. – van LEEUWEN, J. M. J.: "Translocation of polymers in a lattice model", The European Phys. J. E: Soft Matter and Biological Physics 2012, 35, 47.

BIOGRAPHIES

Andrzej Z. Grzybowski received his PhD degree in the field of mathematics from the Institute of Mathematics, Wroclaw University of Technology, Poland, in 1991. His thesis title was "Minimax decisions in some problems of control of stochastic systems". In 2013 he received D.Sc. degree from the Institute of System Engineering and Informatics, University of Pardubice, Czech Republic. His habilitation thesis was entitled "Modeling and simulation in decision making under uncertainty". Since 1983 he works at Czestochowa University of Technology, Poland, where now he is an associate professor at the Institute of Mathematics, Faculty of Mechanical Engineering and Computer Science. His scientific research is focusing on artificial intelligence as well as on formal and Monte Carlo analysis of various problems arising in statistical decision theory.

Zbigniew Domański received the Ph.D and D.Sc. degrees in solid state physics from the Institute of Low Temperature and Structure Research, Polish Academy of Sciences, Wroclaw, Poland, in 1987 and 1997, respectively. From 1988 to 2000 he was a senior researcher with the Institute of Theoretical Physics, University of Lausanne, Switzerland. Now he is a professor with the Faculty of Mechanical Engineering and Computer Science, Czestochowa University of Technology (CUT), Czestochowa, Poland. His research interests range from applied mathematics, statistical mechanics to transport phenomena in nanotechnological systems.

Kamila Bartłomiejczyk graduated from the Faculty of Mechanical Engineering and Computer Science at the Czestochowa University of Technology in 2011 with the degree MSc in informatics. Her thesis title was "E-promotion. How to utilize the Internet network for the promotion?" In 2012 from the same department she received the BSc degree in mathematics - the thesis title was "Polynominal interpolation of real functions". Now she is a PhD student in the field of informatics at the same institution. Her scientific research is focusing on analysis of chain-like structures translocation.