Modeling of Dynamics of Charge Transfer in Biopolymer with Temperature

E.V. Sobolev, N.S. Fialko

Institute of Mathematical Problems of Biology RAS, Institutskaya street 4, Pushchino, 142290, Moscow region, Russia

egor.sobolev@gmail.com, fialka@impb.psn.ru

Abstract. *In Holstein model, the transfer of a charge (non-relativistic quantum particle) along one-dimensional chain of sites and motions of classical sites of the chain are described by the self-consistent ODE system. To estimate the charge mobility value in the chain at finite temperature of thermostat, we use direct modeling - calculation of the set of trajectories and averaging over ensemble,– and apply Cubo formulas. We applied 2 variants of calculation scheme for modeling of charge transfer in biopolymer at temperature prescribed: natural parallelization with the performance of each sample on a separate node, and parallelization of one sample on the multi-core node with shared memory. Programs are written in C using MPI and openMP. Tests were made on the supercomputer mvs-100k of Joint Supercomputer Center RAS. The effectiveness of the natural parallelism is of almost 100%. For calculations of one sample on the 4-core node, speed-up depends on the chain length, with a maximum of 3.3 times.*

Keywords

Holstein model, Langevin equation, ODE system with a large number of variables, MPI, openMP

1 Introduction

At present, attention of researchers attracts possible applications of biological macromolecules in nanobioelectronics, especially DNA, for example, in the development of electronic biochips and using as molecular wires. Therefore, the study of the conducting properties of biopolymers is of interest [1].

The value of the conductivity of the molecular chain can be evaluated, knowing the mobility of the charge and the free charge concentration. The mobility of charge in a quasi-one-dimensional chain can be estimated by applying the method described in [2], i.e.: to calculate the set of samples at a given temperature and find the time-dependence of the charge mean square displacement, than find the diffusion koefficient, which is proportional to mobility.

In computationally, the resource-intensive part of the problem is the calculation of a large number of samples, since the accuracy of the mean is proportional to the square root of the sample amount.

2 Holstein model

The model is based on the Holstein Hamiltonian for discrete chain of sites [3]. In the case of DNA site is a complementary base pair, which is approximately modeled by an oscillator (spring as hydrogen bond). The excess charge (electron or hole) is introduced in the chain. Distribution of the charge affects the movement of classical sites, and site displacement changes the probability of finding the charge on it. The dynamics of the model is described by a system of differential equations, which, in dimensionless form, is

$$i\dot{b}_n = \eta_{n,n-1} b_{n-1} + \eta_{n,n+1} b_{n+1} + \eta_{n,n} b_n + \chi u_n b_n , \quad (1)$$

$$\ddot{u}_n = -\omega^2 u_n - \chi \,|\, b_n \,|^2 - \gamma \dot{u}_n + Z_n(t) , \qquad n = 1,\ldots N. \quad (2)$$

Variables: $b_n(t)$ – the probability amplitude of finding the charge on the n-th site, $u_n(t)$ – displacement of n-th site from its equilibrium, N – number of sites. Parameters: $\eta_{n,n}$ – electron energy on n-th site, $\eta_{n,n\pm1}$ – the transition matrix elements (we consider the approach of the nearest neighbors), ω – frequency of classical oscillator, χ –constant of interaction for quantum and classical subsystems. To simulate the thermostat, the friction term with koeffisient γ and random force $Z_n(t)$ with properties $\langle Z_n(t) \rangle = 0$, $\langle Z_n(t) Z_k(t') \rangle = \mathrm{Const} \cdot k_B \mathrm{T} \gamma \cdot \delta_{kn} \delta(t - t')$, T – temperature, are added in the (2) (Langevin equation). More details about the model, the dimensionless and parameter values, see e.g. [4].

To evaluate the mobility, it is necessary to perform the calculation of the set of samples – the dynamics of the charge distribution from the different initial conditions and with different values of the random force, than to find charge mean-square-displacement $\left\langle \sum_{n=1}^{N} | b_n(t) |^2 \; n^2 \right\rangle$. In computationally, task is to finding the set of trajectories of the system (1), (2).

3 Results

To reduce the computation time in the integration of (1),(2), we applied two variants of program parallelization:
1) natural parallelism with the performance of each sample on a single node, using MPI, and
2) parallelization of one sample, calculated on the multi-core node with shared-memory, using openMP.

Although explicit iterative method use for numerical integration of the ODE system (1), (2), the equations includes explicitly only the nearest neighbors. So we can "devide" the chain into several fragments, which are integrated in one step independently on different cores.

The integration is performed by multi-stage one-step method, synchronization is executed once, at the end of the step. At the stage calculations of the fragments are overlapped (e.g., for 2-stage method we need the initial data from the 4 nearest neighboring sites). According to the test results, it is more profitable to calculate data on the ends of fragments twice on different cores than synchronize at each stage.

Obviously with this method, the longer fragments, the smaller relative part of the additional calculations, and the higher the speed-up T1/Tp (T1 – machine time of sequentional version, Tp – duration of the task, parallelized on p threads). We tested it on the 4-core processors (Intel Xeon E5520, 2.3ГГц), changing the number of threads from 1 to 4. Figure 1 shows the results.

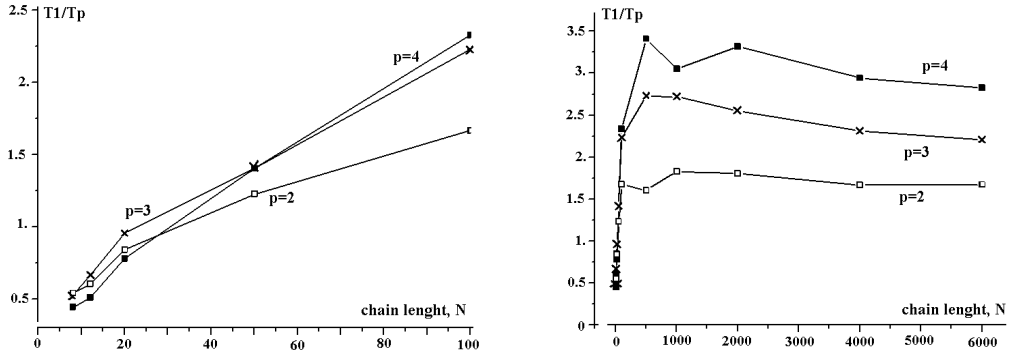

Fig. 1. Dependence T1/Tp on site number N.

For chain length 100 sites, T1/T4 ~ 2.3; for sequences of 1000 sites and more T1/T4 ~ 3. At the right figure, it is seen that the speed-up decreases with chain length increasing. We suppose that this is due to the fact that by increasing the size of data sets, no longer fit in the cache of the processor.

For a fixed number of nodes in the cluster, for equal machin time, we can select the variant based on the type of the problem,– to calculate more short trajectories of N samples (for evaluating of the charge mobility), or to calculate 3 time longer trajectories of N/4 samples (e.g., for modeling the dynamic processes of not-reversible charge transfer, or access to the thermodynamic equilibrium). Choosing variant, also need to take into account the total amount of memory on the node and max task-time for the computing resource.

4 Acknowledgments

The work is supported particullary by Russian Foundation for Basic Research, grant 11-07-00635.

References

[1] Nanobioelectronics – for Electronics, Biology, and Medicine. Eds. Offenhäusser A., Rinaldi R. New York: Springer, 337 p., 2009.
[2] V.D.Lakhno, N.S.Fialko: Hole mobility in a homogeneous nucleotide chain. *JETP Letters*, 78(5): 336–338, 2003.
[3] T. Holstein: Studies of Polaron Motion. *Ann. Phys.*, 8(3): 325–389, 1959.
[4] V.D.Lakhno, N.S.Fialko: Model of charge transfer in DNA (in Russian). In: *The Dynamic Models of the Processes in Cells and Subcellular Structures.* G.Yu. Riznichenko, A.B. Rubin (Eds.), Regular&Chaotics Dinamics, 11–67, 2010.