Research proposal for the DARI campaign 2014, First application:

Coarse grain MD simulations of DNA translocation through nano-pores in graphene-like sheets

Timothée Menais, Stefano Mossa and Arnaud Buhot
SPrAM , INAC, CEA Grenoble, France

Abstract: In this project, we are interested in the new opportunities provided by the use of mono-atomic graphene in the highly active field of the translocation of biopolymers through nano-pores. In the context of the low cost DNA sequencing, we already conducted some preliminary work on narrow pores where the thermal vibrations cannot be neglected any more. This work has shown that the slowing down of the translocation process necessary for DNA sequencing could be met under certain circumstances. We would like to investigate further those circumstances trying a new approach based on functionalized nano-pores.

We ask for the following computational resources :

TGCC Curie : 60 000 hours

Contents

1 Context

The translocation of biomolecules through nano-pores is an active research field concerning both experimental and theoretical aspects. The translocation process consists in a biomolecule crossing a membrane through a nano-hole called nano-pore. This process can either be natural (unbiased) or forced. The first successfull DNA translocation was obtained in 1996 by Kasianowicz et al. [1]. It fostered several single objects manipulation techniques [2] and was followed by several applications, in bio-engineering and drug delivery for example [3]. Other high impact potential applications in biotechnologies and medical techniques can still be expected, like for example a quicker and cheaper method concerning DNA sequencing [4].

Current limitations to the use of DNA translocation as a sequencing tool are both spatial and temporal. In the case of common natural and artificial nano-pores, several nucleotides are simultaneously present within the nano-pore during the translocation. This prohibits the single base sequence resolution [5]. Furthermore, in most experiments, a base spends about 1μs within the pore, but measurement resolution time would require a slowing down of the process leading to an occupation time of around 1ms [6].

Recently, experiments have involved translocations through nano-pores drilled in mono-atomic graphene sheets [7]. This may provide a solution for the thickness issue since the width of the membrane is substantially thinner than a nucleotide. However, the graphene sheet is flexible and therefore can vibrate due to both thermal and elastic fluctuations (see figure 1). This challenges an assumption used in all previous theoretical approaches: The membrane and the nano-pore are considered immobile. Although this assumption was relevant in former experiments with thick membranes, vibration and flexibility cannot be neglected anylonger in the case of graphene sheets and are expected to influence the translocation time of biomolecules.

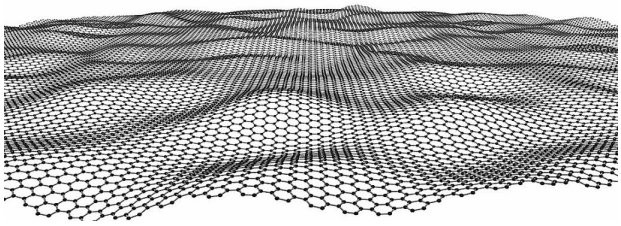

Figure 1: Artist view. The graphene sheet is flexible and therefore can vibrate due to both thermal and elastic fluctuations.

We have conducted a preliminary work presented in section 3 (not published yet) showing that vibrations can indeed modify the translocation process and especially the translocation time. Further numerical simulations are necessary in order to characterize those modifications, motivating the present application for computer calculation time.

2 Basis of the numerical model

We have considered a coarse grain model for the biomolecules, the graphene sheet and their mutual interactions. Our aim is to produce very extended ensembles of translocation events in order to completely caracterize the statistical properties of the translocation time.

The polymer

A minimalistic model for DNA is considered including only the sterical and polymer binding interactions. We distinguish between 3 kind of grains, the phosphates and sugars are forming the polymer backbone and lateral amino-acids are connected to the sugars. No distinction is made between the 4 amino-acids (Adenine, Cythosine, Thymine and Guanine) in this preliminary work.

The sterical interactions are modelled by a Lennard-Jones potential truncated and shifted:

$$U_{EV}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] + \epsilon, \qquad (1)$$

with $U_{EV}(r_{ij}) = 0$ for $r_{ij} > 2^{1/6}\sigma$, r_{ij} being the distance between grains i and j. This potential is purely repulsive and only prevents the grains from interpenetrating at very short distances. Bonds between grains are modelled by a FENE (Finitely Extensible Nonlinear Elastic) anharmonic potential:

$$U_{MH}(r_{ij}) = -15\epsilon \left(\frac{R_0}{\sigma} \right)^2 \ln \left[1 - \left(\frac{r_{ij}}{R_0} \right)^2 \right]. \qquad (2)$$

With this minimalistic model, H binding, stacking and backbone bending are not taken into account. Therefore we are not able to tackle issues including stiffness and helicoidal structure building of the DNA [8]. We have indeed considered that the relevance of these degrees of freedom in the present context is very low compared to the other interactions we have kept. Furthermore, their implementation would have meant an additional heavy computational cost associated at the expense of an accurate translocation sampling.

The values given to the energy scale ϵ, bond length R_0 and L-J potential parameter σ are chosen to match physiological conditions, as explained by M.C. Lynak et al. in [8]. We have already tested the consistency of our polymer model for single polymers in a solvent or grafted to a wall (see appendix).

The graphene sheet

Graphene is a 2D crystal formed by aromatic carbon cycles arranged on a honeycomb lattice. Carbon are thus bonded together both via σ and π bonds characteristic of namely σ and π molecular orbitals, see figure 2. In the graphene, interactions among carbons are made through σ orbitals while elements from the outside will interact with the π electrons.

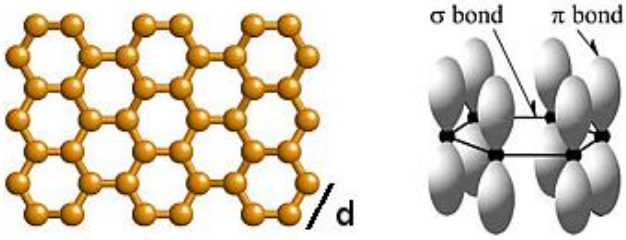

Figure 2: Crystallographic structure and π orbitals of graphene.

In our model, we have kept a consistent size ratio between graphene and DNA, leading to a lattice constant scaling to half the size of a DNA backbone grain. 2D periodic boundary conditions are used to simulate a large graphene sheet which is chosen wide enough to avoid interaction of a fully stretched DNA strand with its own image. This implies considering a large number of carbon grains. Therefore, binding all the carbon atoms together would have been computationally too costly. The flexibility of the graphene sheet as been put aside to focus only on the thermal vibration aspect.

To carry out this investigation, we set up the carbon grains in a harmonic potential centered on their

equilibrium position in the lattice. The carbon interact between one another through σ orbitals and with the polymer grain through π orbitals. The size of the σ and π orbitals was determined using respectively the latice constant and the experimental measure of the thickness of a graphene sheet. Figure 3 is a VMD [9] screenshot of an equilibrated polymer ready to be pulled through the nano-pore drilled in a graphene membrane, the scales are respected.

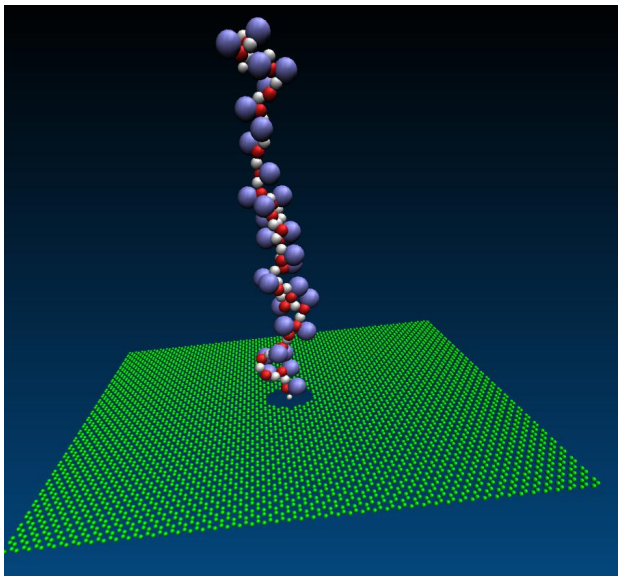

Figure 3: The polymer prior to the translocation. Graphene grains are coloured in green, the polymer backbone in red and white. Lateral grains in purple are bonded to the red ones from the backbone.

Langevin Dynamics

Concerning the time integration, the LAMMPS [10] software solves the Langevin equation:

$$\frac{d\mathbf{r}_n}{dt} = -\frac{1}{\nu_n}\frac{\partial U_n}{\partial \mathbf{r}_n} + \mathbf{g}_n \qquad (3)$$

with $\mathbf{r}_n$ being the position of the n^{th} grain, t the time, U_n the sum of all the potentials applied to the n^{th} grain, ν_n the friction coefficient of the n^{th} grain and $\mathbf{g}_n$ the brownian part in the trajectory. This brownian part has the following properties: a zero average and a variance scaled to the temperature using k_BT:

$$\langle \mathbf{g}_n \rangle = 0 \ , \ \langle \mathbf{g}_n^2 \rangle = 2k_BT. \qquad (4)$$

To fulfil the physiological conditions, the k_BT energie scale is set to 1.5 ϵ [8] in most of our simulations.

Particularity of our approach

Our approach stands aside from many other studies as our model goes beyond the simple linear polymer, mimicking the sterical properties of DNA while remaining simple enough. This allows for numerous translocation simulations to be made in order to obtain enough statistics for the translocation times. This original modelling enables one to investigate the influence of steric interactions especially with a narrow pore, as we will show in the next section.

3 Preliminary results

For a sequencing prospect, one of the issues to be dealt with is the speed of the translocation. Current sensibility in measurement techniques is challenged by a process that is happening too fast in most of the experiments [6]. Using our model we have compared the average translocation time between a large nano-pore fulfilling the hypothesis of most theoretical approaches (one single monomer occupying the pore at a given time and no friction from the pore [11]) and a narrow one. The narrow nano-pore is small enough that the polymer has to tilt to pass through it, the energy landscape is consequently strongly modified. In figure 4, we demonstrate that using such a narrow nano-pore is a way to significantly reduce the translocation time up to a factor 4.

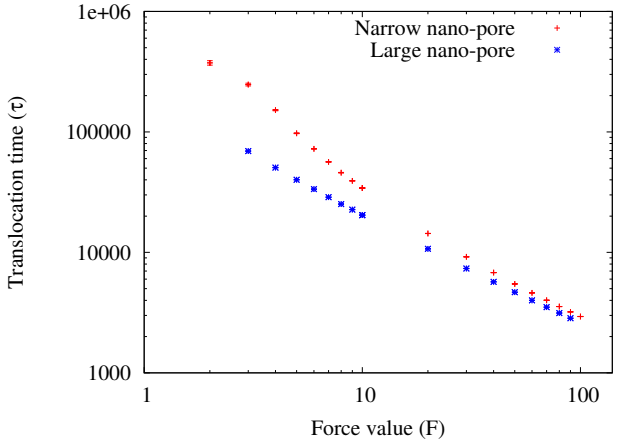

Figure 4: Narrowing the pore slows the translocation process. For strong forces the behaviour of the polymer is not modified, whereas at low forces, the size of the pore strongly influence the translocation.

The result shown in figure 4 were obtained using fix grains for the graphene. We then allowed the grains to move within an harmonic potential and checked out how this would affect the translocation times. Whereas for the large narrow pore no significant effects were found, the small pore allowed to vibrate can strongly affect the translocation.

With the carbon grains set in harmonic potential traps we have explored several configurations, vary-

ing the strength of traps, the mass of the grain and their temperature. Figure 5 shows some of the experiments we have conducted.

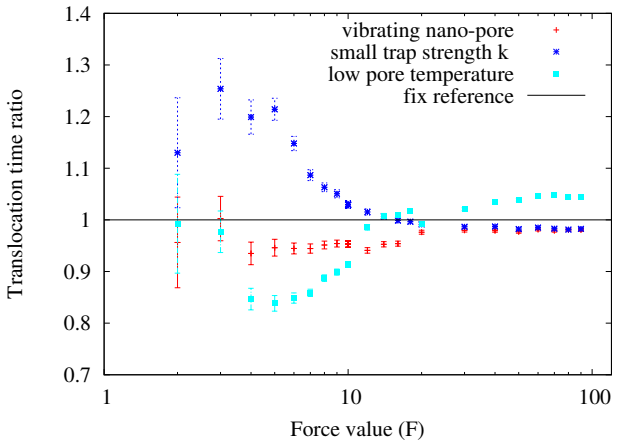

Figure 5: Ratio of the mean translocation time in different experiments with a vibrating graphene sheet compared to a fixed reference. The introduction of vibrations modifies the average translocation times observed.

During the experiments, our model polymer translocates gradually since each lateral grain is temporary blocked at the entry of the pore (see appendix). This gradual process could be affected by the frequencies of the harmonic traps ν funciton of the strength k and the mass m linked by equation 5.

$$\nu = \sqrt{\frac{k}{m}} \tag{5}$$

The results shown in Figure 5 suggest that ν and the pore temperature could, depending on their values, either accelerate or slow down the translocation process.

4 Perspectives

This preliminary work obviously not complete already suggest interesting perspectives for the slowing down of the translocation. Despite having a possible influence, the vibrations alone will not be sufficient to reduce the translocation time by three orders of magnitude which is the goal for possible effective DNA sequencing. In this project, we propose to investigate another way to significantly reduce the translocation time using functionalized nano-pores.

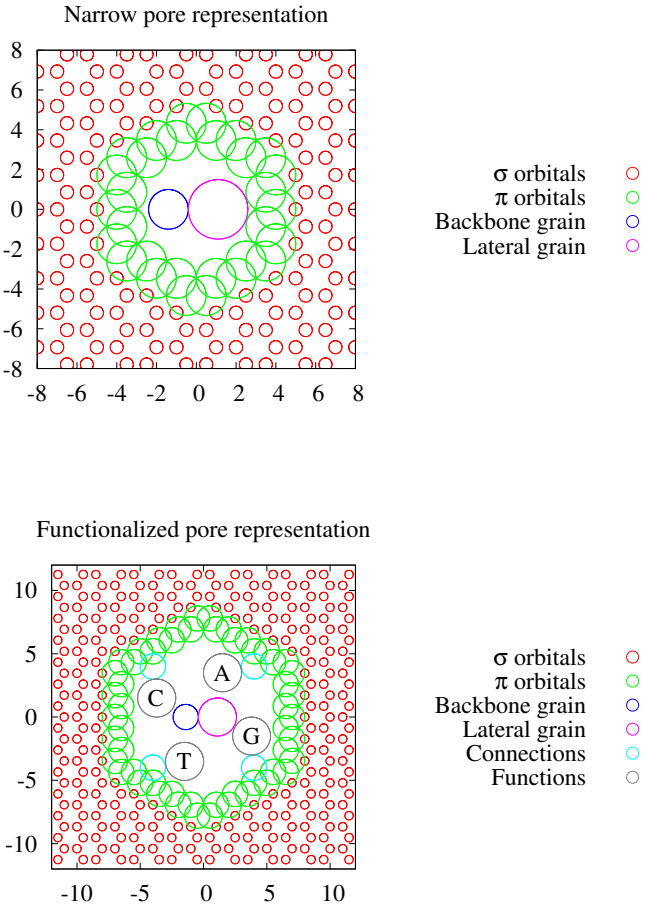

Figure 6: Top: Narrow nano-pore used in the preliminary study. Bottom: Functionalized nano-pore.

In a previous collaboration with an experimental team working on micro-pores, functionalized pores were considered and shown to affect the translocation of cells and nano-objects [12]. Functionalization could be a smart way to meet the slow process criteria also in nano-pores. We propose to functionalize the pore by grafting the 4 different amino-acids within the pore, as illustrated in figure 6. The H-binding interactions of those amino-acids with their counterpart amino-acid in the polymer should affect strongly the translocation speed. Such interactions are deemed to highly increase the energy barrier that the monomers will have to cross. This could thus decrease the translocation time in a drastic way, hopefully reducing it by a few orders of magnitude.

5 The computing tool

The model we have described previously is implemented in a Molecular Dynamics solver, LAMMPS [10] that solves the Langevin equation.

This software is mentioned on the list of the possible requirements of the website of the DARI campaign and should already be available on the computers (or easily implementable). It is highly efficient and its use is wide spread among the computing modelling community.

The most accurate decription of the software is given by its developers: LAMMPS is a classical molecular dynamics code that models an ensemble of particles in a liquid, solid, or gaseous state. It can model atomic, polymeric, biological, metallic, granular, and coarse-grained systems using a variety of force fields and boundary conditions. In the most general sense, LAMMPS integrates Newton's equations of motion for collections of atoms, molecules, or macroscopic particles that interact via short- or long-range forces with a variety of initial and/or boundary conditions. For computational efficiency LAMMPS uses neighbor lists to keep track of nearby particles. The lists are optimized for systems with particles that are repulsive at short distances, so that the local density of particles never becomes too large. On parallel machines, LAMMPS uses spatial-decomposition techniques to partition the simulation domain into small 3d sub-domains, one of which is assigned to each processor. Processors communicate and store "ghost" atom information for atoms that border their sub-domain.

LAMMPS has already been used by the PhD student working on the project (see preliminary work section and the internship report in appendix) and one of the supervisor already published several papers using it [13, 14].

6 Computational cost

In our simulations, we first need independent trajectories to ensure significant statistics. We wrote a C routine to provide LAMMPS with the nano-pore in the graphene lattice and the polymer with an end fixed in the middle of the pore. The system is then left to equilibrate and restart points are saved. The restart points are uncorrelated (see correlation times in appendix) initial conditions to begin the translocations. It takes around 80 hours of cpu time to generate a thousand of those initial conditions. New initial conditions are needed whenever the vibrational parameters are modified.

While carrying out the preliminary work, we have estimated the cpu time it has needed. We worked on a small cluster providing 48 cores on 6 identical nodes. The 8 processors (Quad-Core AMD Opteron(tm) Processor 2354, 1100 MHz, cache 512 KB) of each node were used to perform the jobs.

Depending on the pulling force applied, the physical time of a translocation is modified, so is the computing time. We report in the following tabular the average computational cost necessary for a single translocation event within a typical force range applied to a 16 monomers long polymer:

Force applied	Average cpu time (s)
0.4	714.64
5	971.18
10	361.12
26	218.08
40	81.55

For small forces, the translocation process is not necessary continuous and many trajectories are aborted when the polymer escapes the nano-pore and does not cross the membrane. This leads to a weaker statistic for small forces and a peak in cpu time for the first force values being high enough to allow an almost 100% translocation rate.

For a 16 monomers long polymer we estimate that 1000 hours (2000 hours on our cluster, the CURIE processors being twice as performant as ours) of cpu time are needed to explore a relevant force range with enough statistics at a given vibrational state.

We would like in the future to test for longer polymers so as to analyze the size behaviour on translocation time. Considering this possibility we tested the computing time required for longer polymers:

Polymer length / force	Average cpu time (s)
16 (f=10)	361.12
16 (f=40)	81.55
32 (f=10)	1187.92
32 (f=20)	552.33
64 (f=10)	5541.49
64 (f=40)	1177.68

A relevant way to compare the previous data is to consider the force per monomer e.g. compare a 64 monomer long polymer pulled with a force of 40 to a 16 long one pulled with a force of $F = 10$. Exploring a vibrational state for polymer of length 32 and 64 are respectively expected to cost around 1700 and 3000 hours of cpu time.

We plan to devote between 15 000 and 20 000 hours of cpu time towards this vibrational issue.

For the pore functionalisation project, we do not have data yet. On the first hand we are expecting a slow down of the translocation process that would lead to an increase in the computational cost, on the other hand the number of vibrating elements can be highly reduced. We are expecting the computational cost to be of the same order of magnitude as in the previous study.

We hence estimate that this project needs 60 000 hours of computing time on the TGCC Curie platform.

References [12, 13, 14] are work from the team.

References

[1] J. Kasianowicz, E. Brandin, D. Branton, and D. Deamer, "Characterization of individual polynucleotide molecules using a membranechannel," *Proceedings of the National Academy of Sciences*, vol. 93, no. 24, pp. 13770–13773, 1996.

[2] U. F. Keyser, "Controlling molecular transport through nanopores," *Journal of The Royal Society Interface*, vol. 8, Oct 2011.

[3] A. Meller, "Dynamics of polynucleotide transport through nanometre-scale pores," *Journal of Physics: Condensed Matter*, vol. 15, no. 17, p. R581, 2003.

[4] H. Bayley, "Nanotechnology: Holes with an edge," *Nature*, vol. 467, p. 164, Sep 2010.

[5] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, "Dna translocation through graphene nanopores," *Nano Letters*, vol. 10, no. 8, pp. 3163–3167, 2010.

[6] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss, "The potential and challenges of nanopore sequencing," *Nat Biotech*, vol. 26, pp. 1146–1153, Oct 2008.

[7] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker, "Dna translocation through graphene nanopores," *Nano Letters*, vol. 10, no. 8, pp. 3163–3167, 2010.

[8] M. C. Linak, R. Tourdot, and K. D. Dorfman, "Moving beyond watson–crick models of coarse grained dna dynamics," *The Journal of Chemical Physics*, vol. 135, no. 20, p. 205102, 2011.

[9] W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[10] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *Journal of Computational Physics*, vol. 117, no. 1, pp. 1 – 19, 1995.

[11] A. Milchev, "Single-polymer dynamics under constraints: scaling theory and computer experiment," *Journal of Physics: Condensed Matter*, vol. 23, no. 10, p. 103101, 2011.

[12] J. Liu, P. Pham, V. Haguet, F. Sauter-Starace, L. Leroy, A. Roget, E. Descamps, A. Bouchet, A. Buhot, P. Mailley, and T. Livache, "Polarization-induced local pore-wall functionalization for biosensing: From micropore to nanopore," *Analytical Chemistry*, vol. 84, no. 7, pp. 3254–3261, 2012.

[13] D. Damasceno Borges, A. A. Franco, K. Malek, G. Gebel, and S. Mossa, "Inhomogeneous transport in model hydrated polymer electrolyte supported ultrathin films," *ACS Nano*, vol. 7, no. 8, pp. 6767–6773, 2013.

[14] G. Monaco and S. Mossa, "Anomalous properties of the acoustic excitations in glasses on the mesoscopic length scale," *Proceedings of the National Academy of Sciences*, vol. 106, no. 40, pp. 16907–16912, 2009.