Minimal model of transcriptional elongation processes with pauses

Jingkui Wang, Benjamin Pfeuty, Quentin Thommen, M. Carmen Romano, Marc Lefranc

Minimal model of transcriptional elongation processes with pauses

Jingkui Wang,[1] Benjamin Pfeuty,[1] Quentin Thommen,[1] M. Carmen Romano,[2,3] and Marc Lefranc[1]

[1] *Laboratoire de Physique des Lasers, Atomes, et Molécules, Université Lille 1, CNRS UMR 8523, ,F-59655 Villeneuve d'Ascq, France*

[2] *SUPA, Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom*

[3] *Institute of Medical Sciences, Foresterhill, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom*

(Dated: November 18, 2014)

Fundamental biological processes such as transcription and translation, where a genetic sequence is sequentially read by a macromolecule, have been well described by a classical model of non-equilibrium statistical physics, the totally asymmetric exclusion principle (TASEP). This model describes particles hopping between sites of a one-dimensional lattice, with the particle current determining the transcription or translation rate. An open problem is how to analyze a TASEP where particles can pause randomly, as has been observed during transcription. In this work, we report that surprisingly, a simple mean-field model predicts well the particle current for all values of the average pause duration, using a simple description of blocking behind paused particles.

PACS numbers: 87.16.Uv, 87.16.aj, 02.50.Ey, 87.16.Yc
Keywords: Transcription, Translation, TASEP, ubiquitous pauses

Transcription is a fundamental biological process in which a genetic sequence carried by DNA is copied to a messenger RNA molecule (mRNA), which is subsequently translated to protein molecules [1]. The transcription rate is one of the most important quantities which contribute to controlling protein levels in the cell. Transcription is described as a sequential and stochastic elongation of a DNA polymer through enzymes called RNA polymerases (RNAPs). An RNAP binds to a specific DNA region adjacent to the gene, called promoter, and slides forwards until it recognises the beginning of the gene. Then, as the RNAP moves forward, it adds a nucleotide to the growing RNA chain complementary to its current position on the DNA template, until it finds the stop or terminator signal, when the produced mRNA is released [1]. Importantly, as soon as the first RNAP moves a sufficient number of nucleotides downstream the gene, a new RNAP can bind to the gene and start a new round of transcription. Hence, transcription can be thought of a production line with several RNAPs bound to the same gene. Importantly, RNAPs cannot overtake each other, and hence transcription rate is expected to be highly dependent on the density and interactions among RNAPs,all the more as RNAPs can temporarily and stochastically stop elongation [2, 3]. These stochastic pauses generate complex traffic dynamics: a paused RNAP in a dense RNAP traffic may cause a traffic jam, forcing multiple trailing RNAPs to stop as well, hence significantly slowing down the overall transcription rate and inducing bursty elongation rates [4–7].

A powerful model to study elongation processes is the totally asymmetric simple exclusion process (TASEP), a paradigmatic model in non-equilibrium statistical physics [8, 10, 11]. It has been successfully applied to describe a wide variety of transport processes [12, 13], especially in biology, such as in the cases of transcription, translation and molecular motors [4–7, 14–25]. A TASEP model describes the dynamics of particles hopping from site to site along a one-dimensional lattice, subject to the condition that the next site is empty.

In this Letter we study the effect of stochastic pauses of the driven particles on the overall traffic dynamics. These pauses describe reversible conformational changes of the particles, rendering them inactive for certain time interval, i.e., particles can pause for exponentially-distributed times before resuming elongation [5]. This model is motivated by the so-called ubiquitous pauses experimentally observed in RNAPs during transcription [2, 3], but is more generally applicable to other driven diffusion processes where the particles can undergo a reversible internal kinetic cycle proceeding in parallel with the hopping dynamics. In spite of the general relevance of this process, especially in biology, it has been little studied theoretically, probably because it was believed that the strong correlations between particles induced by jamming complicate significantly the analysis. We show here that, surprisingly, a simple approximation leads to an effective mean-field model which captures quantitatively the effect of the pausing process on elongation dynamics for short as well as for long pauses, and in particular the dramatic decrease of particle current with pause duration. A simple but key ingredient is to consider that particles colliding with a paused particle become blocked and effectively behave themselves as paused particles.

TASEP model with pauses. We study an exclusion process where N_p particles move unidirectionally on a one-dimensional lattice of size N (Fig. 1). For simplicity, we consider periodic boundaries, but we briefly discuss at the end how our results can be extended to the open boundary case. Moreover, particles can switch between an active state and a paused state, with respective probabilities per unit time f and $1/\tau$, following the notation

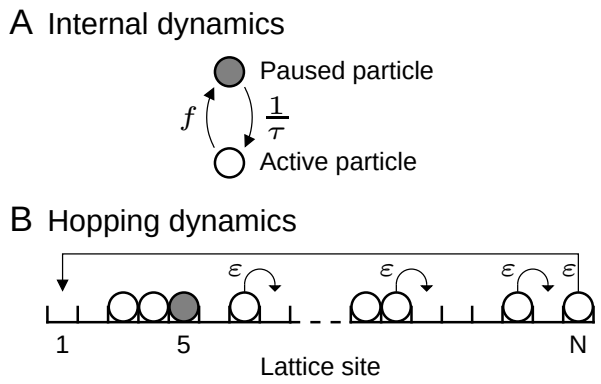

FIG. 1. TASEP model with pausing. (A) Particles can transition from an active state (in white) to a paused state (in grey) and back, with respective rates f and $1/\tau$. (B) Active particles can hop to the next site if it is empty (exclusion process), with unit-time probability ϵ. Paused particles do not move.

used in [5]. If the next site is empty, an active particle moves forward with probability ϵ per unit time, whereas a paused particle remains immobile in all cases (Fig. 1). We restrict our analysis to the thermodynamic limit in which $N_p, N \to +\infty$, while keeping constant the average particle density $\rho = N_p/N$.

In the standard mean-field approximation, where correlations between occupation of neighbouring sites are neglected, the probabilities a_i and p_i of a particle at site i being active and paused, respectively, are given by the following equations

$$\frac{da_i}{dt} = \epsilon a_{i-1}(1-a_i) + \frac{1}{\tau}p_i - \epsilon a_i(1-a_{i+1}) - f a_i$$
$$\frac{dp_i}{dt} = f a_i - \frac{1}{\tau}p_i,$$

where $a_i + p_i = \rho_i$, the average occupancy of site i. Given the system's symmetry, $a_i = a$, $p_i = p$, and $\rho_i = \rho\ \forall i$. Moreover, considering the steady state, the average fraction $\phi = p/\rho$ of paused particles is given by

$$\phi = \frac{f\tau}{1+f\tau}. \qquad (1)$$

The average densities of active and paused particles are then equal to $a = \rho(1-\phi)$ and $p = \rho\phi$, respectively. The main quantity of interest to compute is the particle current J, i.e. the average number of particles arriving per unit time on any given site, which biologically corresponds to the transcription rate. For the standard TASEP without pauses, the particle current in the mean-field approximation is given by $J_0 = \epsilon\rho(1-\rho)$, which becomes exact in the thermodynamic limit [8].

Mean-field approximation for short pauses. We first consider the case of short pauses, when the internal switching dynamics is faster than the hopping dynamics

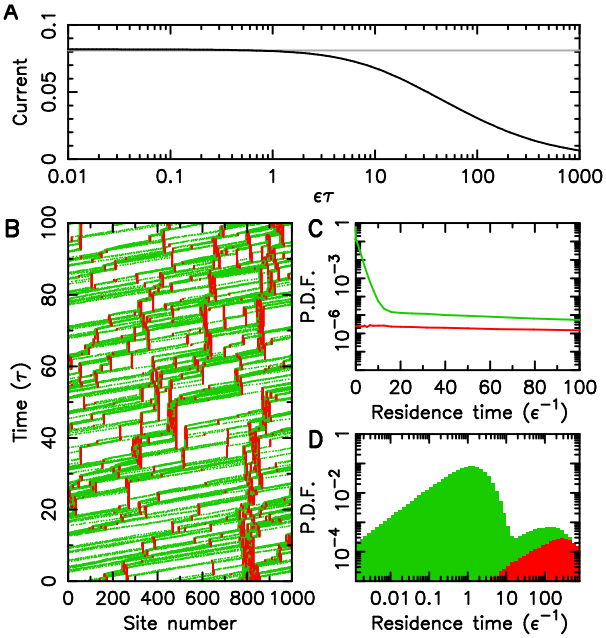

FIG. 2. Occurrence of traffic jams on a TASEP with long pauses. (A) Particle current J as a function of the mean pause duration $\epsilon\tau$: numerical estimate in black (with error bar) compared to the mean field value J_1 given by Eq. (2) (grey line). Computation is performed with 1000 particles on a lattice of 10000 sites keeping $f\tau = 0.1$ fixed. (B) Temporal evolution of the particle distribution on the lattice when $\epsilon\tau = 100$. Sites occupied by an active (resp., paused) particles are shown in green (resp., red). (C)-(D) Probability density functions of the residence time (C) and of the logarithm of residence time (D) for particles that remained active at any time (in green) and for particles that have paused at least once (in red), also for $\epsilon\tau = 100$.

along the lattice ($\epsilon\tau \ll 1$). The internal state of a particle at the time when it is eligible for advancing can then be considered as random, with a probability $(1-\phi)$ to be in the active state. Thus, the probability that a given site contains a particle ready to advance is $\rho(1-\phi)$, so that the expression of the current is

$$J_1 = \epsilon\rho(1-\phi)(1-\rho) = \epsilon\frac{\rho(1-\rho)}{1+f\tau} = \frac{J_0}{1+f\tau}. \qquad (2)$$

Observation of clustering in numerical simulations. To assess the validity of expression (2), we used stochastic simulations based on the Gibson-Bruck algorithm [9] to compute the particle current J for a wide range of mean pause durations τ, keeping $f\tau$ fixed at 0.1, which is a biologically relevant value [5]. Under this constraint, the fraction ϕ of paused particles does not change with τ ($\phi \sim 0.091$) so that J should remain constant according to (2). The result of these numerical simulations is shown in Fig. 2A. It can be seen that although expression (2) is valid when $\epsilon\tau \lesssim 1$, it does not describe correctly the TASEP dynamics when pauses are too long. As the mean pause duration τ increases gradually beyond

the characteristic time $1/\epsilon$, the particle current decreases monotonously and tends to zero for very long pauses.

For large τ, many particles remain immobile long enough to block trailing particles thus creating local inhomogeneities in the form of traffic jams. This is illustrated in Fig. 2B where the spatio-temporal evolution of particle density is represented for the biologically relevant case $\epsilon\tau = 100$ [5], showing the occurrence of shocks, materialized by clusters of particles remaining stopped in a given location for a significant time. Because of the correlations between particles so induced, the assumptions leading to the mean-field expression (2) fail. Indeed, there is then a sizeable fraction of active particles which should contribute to current according to this approximation but actually do not, because they are blocked behind a paused particle.

A natural way to take this effect into account is to consider a third state besides active and paused particles, comprising active particles blocked behind a paused particle. The relevance of such a state is easily evidenced by studying the asymptotic distribution of particle residence times at a site of the lattice. Figure 2C shows the probability distribution function (PDF) of these waiting times, for particles having paused at least once or not at all while at the site. Since elongation times are exponentially distributed, the PDF of the logarithm of residence time, shown in Fig. 2D, is more revealing and shows that among those particles that have never paused while at the site, there is a significant fraction whose waiting time statistics are very similar to those of particles which have paused at least once. These particles are active yet behave liked paused particles and must be distinguished. Hence, it is convenient to introduce a third state.

Effective mean-field model of the TASEP with long pauses. To compute the particle current, we need to estimate the fraction of particles actually contributing to it, namely those which are both active and not blocked behind a paused particle. The size of this fraction depends on the collective state determined by the TASEP parameters ρ, ϕ and τ. To this aim, we extend the TASEP model of Fig. 1 by adding a blocked state. Active particles can enter this blocked state by colliding with a paused particle (directly or indirectly via particles blocked behind a paused particle). Since the density of paused particles is $\rho\phi$ and the advance rate is ϵ, it follows that in a mean-field type approximation, an active particle has unit-time probability $\kappa = \epsilon\rho\phi$ of becoming blocked.

Particles in a blocked state return to active state after the blocking pause ends. Neglecting the transient following unblocking of a cluster, it is natural to assume that as soon as a particle resumes from pause, all blocked particles behind it also return to the active state. Under this assumption, the unit-time probability of returning to active state is the same for blocked and paused particles, namely $1/\tau$. Remarkably, this allows us to simplify the model by grouping paused and blocked particles in a single paused/blocked state. Active particles enter the paused/blocked state with a rate $\tilde{f} = f + \kappa$ which is the sum of the pausing and blocking rates, and leave it with rate $1/\tau$. This amounts to recasting the three-state TASEP into the two-state TASEP of Fig. 1, but with a modified pausing rate $\tilde{f}$. The particle current is then simply given by

$$J_2 = \frac{\epsilon\,\rho\,(1-\rho)}{1+\tilde{f}\tau} = \frac{\epsilon\,\rho\,(1-\rho)}{1+(f+\epsilon\rho\phi)\tau}. \tag{3}$$

Remarkably, this expression can be rewritten as

$$J_2 = \frac{J_1}{1+\gamma\tau} = \frac{J_0}{(1+f\tau)(1+\gamma\tau)} \tag{4}$$

where

$$\gamma = \epsilon\rho\phi\ (1-\phi) = \kappa\ (1-\phi) \tag{5}$$

is the probability per unit time that a non-blocked particle, paused or not, becomes blocked ($1-\phi$ is the probability of not being in pause).

In spite of their simplicity, Eqs. (4) and (5) predict remarkably well the particle current J observed in numerical simulations of the TASEP model of Fig. 1. This is shown in Fig. 3, where we have plotted J/J_1 to better visualize the correction brought by our analysis to the short-pause mean-field model (2). In this numerical test, the density ρ and the fraction of paused particles ϕ are scanned independently between 0.1 and 0.95. For each pair of values of ρ and ϕ, a sequence of pause durations τ are chosen so that the values of $\gamma\tau$ are regularly spaced on a logarithmic scale between 10^{-3} and 10^{3}, and the particle current is computed for each of these. Besides assessing the validity of (4) in the entire parameter space (ρ, f, τ), Fig. 3, together with Fig. 2, also indicates that the key parameters controlling the TASEP dynamics are the no-pause current $J_0 = \epsilon\rho(1-\rho)$, $f\tau$ and $\gamma\tau$. Moreover, expression (4) describes accurately the long-pause behavior of the particle current, which is

$$J \approx J_\infty = \frac{1-\rho}{\tau\phi} \quad (\tau \to +\infty). \tag{6}$$

Simple derivation of the asymptotic long-pause current The asymptotic expression (6) can also be obtained by considering the dynamics of clusters of particles blocked behind paused particles. For large τ, particles are most of the time either paused or blocked in such clusters. Then, the only contribution to the current arises when the paused head of a cluster becomes active again, and advances towards the next cluster together with its trailing particles. There are on average $n_c = N\rho\phi$ clusters, whose lengths are distributed around an average of $l = 1/\phi$ sites. The average empty space between two clusters is $d = (1-\rho)/\rho\phi$, which is the total empty space $N(1-\rho)$ divided by n_c. Assume that the lengths of a cluster and of the empty space in front of it are uncorrelated.

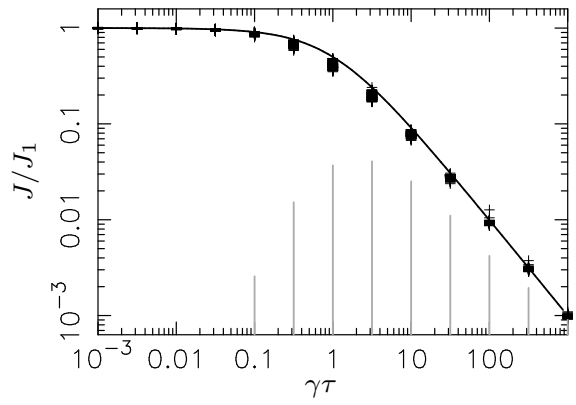

FIG. 3. Comparison between current predicted by Eqs. (4) and (5) (solid line) and numerical estimates (crosses) over a wide parameter range. Grey bars represent the averaged distance $\langle|\tilde{J}_i - J_2|\rangle_i$ between theoretical J_2 and numerical estimates $\tilde{J}_i$.

When a cluster head resumes to active state, which occurs with a probability per unit time of $1/\tau$, there are on average l particles advancing by d sites. The asymptotic current, defined as the average number of particle hops per site and per unit time, is then given by $n_c l d/N\tau$, which simplifies to (6). Presumably, this expression can be derived rigorously in the long-pause limit.

We indeed found that expression (4) matches very accurately numerical simulations in the short-pause and in the long-pause limits, and remains precise when the time scales for advancing and for relaxing from pause are similar (Fig. 3), typically for $1 < \epsilon\tau < 10$. Improving agreement in this intermediate region would require to take statistical fluctuations into account, for example with a master equation approach.

A natural question is whether our analysis can be extended to the more biogically realistic open boundary conditions. Particle density is then no longer a parameter but depends on entry and exit rates (associated to transcription initiation and termination, respectively). Numerical simulations of the open boundary TASEP show that relation (4) between the particle current and the average particle density agrees is still verified. We therefore believe that our analysis is also valid in this context.

In conclusion, we have studied the effect of stochastic pauses on a driven diffusion process, motivated by the fundamental biological process of transcription. This model is however more general, since it describes the effect of an internal reversible kinetic cycle of the driven particles on the overall traffic dynamics. We have shown that stochastic pauses lead to complex traffic dynamics characterised by the creation of local inhomogeneities in the form of shocks. Despite the emergent complex behaviour, we have proposed an effective mean-field description of the process that reproduces the numerical results quite accurately. Taking into account the active particles blocked behind a paused particle, which thus do not contribute to particle current J, allowed us to derive a simple expression for J. This expression is presumably exact for short pauses as well as for long pauses and provides a good approximation between these two limiting cases. Together with the identification of the key parameters controlling the current ($\epsilon\rho(1-\rho)$, $f\tau$ and $\gamma\tau$), this surprising result should be a useful guide for subsequent analytical treatments of the TASEP with pauses, both with closed or open boundary conditions.

This work complements other descriptions of driven diffusion processes in which particles are endowed with an internal irreversible kinetic cycle, such as in translation [15, 20, 28]. Interestingly, also there long-range correlations are generated by the particle stepping cycle, but a simple mean-field approach has been shown to accurately describe the biologically relevant regime of long pauses compared to translocation. One important difference, however, is that in the model considered here [4, 5], the internal cycle proceeds independently of translocation. Thus, particles can perform several consecutive elongation steps without becoming inactive or undergo several internal cycles without advancing. In contrast, in the models presented in [15, 20, 28], particles can elongate only after completing their internal kinetic cycle.

The process studied in this work represents a minimal model of transcription elongation, and it highlights the crucial effect that the ubiquitous pauses experimentally measured for RNAPs can have on the overall transcription dynamics. State of the art techniques allow single molecule transcription experiments to be performed [2, 3]. Our results provide insight into the key pauses' effects and can therefore help interpreting these experimental data. Next research steps in this direction include the study of backtracking pauses [6, 7] within this framework and the coupling between transcription and translation in bacteria [29].

Acknowledgments. M.C.R. is funded by Biotechnology and Biological Sciences Research Council (BB/F00513/X1) and SULSA, and acknowledges a Visiting Professorship at Université Lille 1 in June 2012. This work has been partially supported by Ministry of Higher Education and Research, Nord-Pas de Calais Regional Council and FEDER through the Contrat de Projets État-Région (CPER) 2007–2013, as well as by LABEX CEMPI (ANR-11-LABX-0007) operated by the French National Research Agency (ANR).

[1] Alberts, B. *et al.*, *Molecular Biology of the Cell* (Garland Science, 2008).

[2] Davenport R.J., Wuite G.J.L., Landick R., and Bustamante C., Single molecule study of transcriptional pausing and arrest by E coli RNA polymerase. Science

287:2497-2500 (2000).
[3] Neuman K.C., Abbondanzieri E.A., Landick R., Gelles J., and Block S.M., Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell, 115:437-447 (2003)
[4] Dobrzynski M, Bruggeman F.J., Elongation dynamics shape bursty transcription and translation. Proc Natl Acad Sci USA, 106:2583 (2009).
[5] Klumpp S., Hwa T., Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. Proc Natl Acad Sci U S A. 105(47):18159-64 (2008).
[6] Klumpp S., Pausing and backtracking in transcription under dense traffic conditions, J. Stat. Phys. 142, 1252–1267 (2011).
[7] Sahoo M. and Klumpp S., bactracking dynamics of RNA polymerases: pausing and error correction, J. Phys.: Condens. Matter 25, 374104 (2013).
[8] Derrida, B. and Domany, E. and Mukamel, D., An exact solution of a one-dimensional asymmetric exclusion model with open boundaries, J. Stat. Phys. 69, 667-687 (1992).
[9] Gibson, M. A. and Bruck, J., Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels, J. Phys. Chem. A 104 (9), 1876–1889 (2000).
[10] Derrida, B., Evans M.R., Hakim, V. and Pasquier V., Exact solution of a 1D asymmetric exclusion model using a matrix formulation, J. Phys A Math. Gen. 26, 1493-1517 (1993).
[11] Gorissen, M., Lazarescu, A., Mallick, K., Vanderzande, C., Current statistics of the asymmetric simple exclusion process with open boundaries. Physical Review Letters 109, 170601 (2012).
[12] O'Loan, O. J., Evans, M. R., Cates, M. E., Jamming transition in a homogeneous one-dimensional system: The bus route model, Phys. Rev E 72, 1404-1418 (1998).
[13] Neri, I., Kern, N. and Parmeggiani, A., Totally Asymmetric Simple Exclusion Process on Networks, Phys. Rev. Lett. 107, 068702 (2011).
[14] Tripathi T., and Chowdhury D., Transcriptional bursts: A unified model of machines and mechanisms, EPL 84, 68004 (2008).
[15] Tripathi T., and Chowdhury D., Interacting RNA polymerase motors on a DNA track: Effects of traffic congestion and intrinsic noise on RNA synthesis, Phys. Rev. E 77, 011921 (2008).
[16] Tripathi T., Schütz G. M., and Chowdhury D., RNA polymerase motors: dwell time distribution, velocity and dynamical phases, J. Stat. Mech. 2009, P08018 (2009).
[17] Romano M.C., Thiel M., Stansfield I., Grebogi C., Queueing phase transition: theory of translation. Phys Rev Lett. 102, 198104 (2009).
[18] Chou T. and Lakatos G., Clustered Bottlenecks in mRNA Translation and Protein Synthesis Phys. Rev. Lett. 93, 198101 (2004).
[19] Shaw L.B., Zia R.K.P. and Lee K.H., Totally asymmetric exclusion process with extended objects: A model for protein synthesis, Phys Rev E 68, 021910 (2003).
[20] Ciandrini, L. , Stansfield, I., Romano, M. C., Role of the particles stepping cycle in an asymmetric exclusion process: A model of mRNA translation, Phys. Rev. E 81, 051904 (2010).
[21] Cook J. and Zia R.K.P. , Feedback and fluctuations in a totally asymmetric simple exclusion process with finite resources. J. Stat. Mech. Theor. Exp., P02012 (2009).
[22] Brackley, C.A. , Romano M.C., Grebogi, C. and Thiel, M., Limited Resources in a Driven Diffusion Process, Phys. Rev. Lett. 105, 078102 (2010).
[23] Greulich, P., Ciandrini, L., Allen R.J., and Romano M.C., Mixed population of competing totally asymmetric simple exclusion processes with a shared reservoir of particles, Phys. Rev. E 85, 011142 (2012).
[24] Parmeggiani A., Franosch T., Frey E., Totally asymmetric simple exclusion process with Langmuir kinetics, Physical Review E 70, 046101 (2004).
[25] Chowdhury, D. , Stochastic mechano-chemical kinetics of molecular motors: A multidisciplinary enterprise from a physicist's perspective, Physics Reports 529, 1-197 (2013).
[26] Chou T. and Lakatos G., Totally asymmetric exclusion processes with particles of arbitrary size. J Phys A: Math Gen 36, 2027 (2003).
[27] Gorissen, M., Vanderzande, C., Finite size scaling of current fluctuations in the totally asymmetric exclusion process. Journal of Physics A: Math. Theor. 44, 115005 (2011). .
[28] Klumpp, S., Chai, S. and Lipowsky, R., Effects of the chemomechanical stepping cycle on the traffic of molecular motors, Phys. Rev. E, vol. 78, p. 041909 (2008).
[29] Proshkin, S., Rahmouni, R., Mironov A., and Nudler, E., Cooperation between translating ribosomes and RNA polymerase in transcription elongation, Science 328, 504-508 (2010).