

Stationary RNA polymerase fluctuations during transcription elongation

V. Belitsky*

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, CEP 05508-090 São Paulo, São Paulo, Brazil

G. M. Schütz†

Institute of Complex Systems II, Theoretical Soft Matter and Biophysics, Forschungszentrum Jülich, 52425 Jülich, Germany

(Received 27 September 2018; published 7 January 2019)

We study fluctuation effects of nonsteric molecular interactions between RNA polymerase (RNAP) motors that move simultaneously on the same DNA track during transcription elongation. Based on a stochastic model that allows for the exact analytical computation of the stationary distribution of RNAPs as a function of their density, interaction strength, nucleoside triphosphate concentration, and rate of pyrophosphate release we predict an almost geometric headway distribution of subsequent RNAP transcribing on the same DNA segment. The localization length which characterizes the decay of the headway distribution depends directly only the average density of RNAP and the interaction strength, but not on specific single-RNAP properties. Density correlations are predicted to decay exponentially with the distance (in units of DNA base pairs), with a correlation length that is significantly shorter than the localization length.

DOI: [10.1103/PhysRevE.99.012405](https://doi.org/10.1103/PhysRevE.99.012405)**I. INTRODUCTION**

DNA transcription is the ubiquitous process that transcribes the information coded in the base pair sequence of DNA into an RNA. The molecular “engine” that performs this task is RNA polymerase (RNAP), which synthesizes an RNA as determined by the base-pair sequence of the DNA [1,2]. To this end, the RNAP locally creates the so-called transcription bubble by unzipping the two DNA strands as it progresses on one of the two single DNA strands. The RNA is polymerized by the RNAP by the addition of nucleotides as the RNAP moves along the DNA, thus forming the so-called transcription elongation complex (TEC). Each translocation from one DNA base pair to the next consists of a cycle of molecular reorganizations of the TEC whose main steps are nucleotriphosphate (NTP) binding, NTP hydrolysis and release of pyrophosphate (PP_i), RNA chain elongation and forward translocation along the DNA template. Thus RNAP plays a central role in gene expression and also as therapeutic drug target [3].

The intrinsically stochastic translocation of a *single* RNAP has been studied in great detail from different perspectives and using different approaches, both theoretical and experimental [2,4–7]. We follow the authors of [4,8–10] and describe the kinetics of single-RNAP elongation by a biased random walk of RNAP along the DNA with a step length of one base pair for each translocation. This approach allows for describing the stochasticity of the step cycle of a single RNAP due to thermal fluctuations. In the setting studied by Wang [4], the slowest processes that mostly determine the average speed of an RNAP and thus the RNA elongation rate are the release of PP_i

and the forward step of the RNAP along the DNA template by one base pair (bp). This reduced description ignores the sequence dependence of the translocation kinetics [11,12], but nevertheless accounts rather well for various experimentally established features of the kinetics of a single RNAP, and is adequate as a starting point for the purposes of the present work in which our main interest is the impact of interactions between RNAP on the fluctuations which lead to fluctuations in the overall rate of elongation, which is proportional to the stationary flux of RNAP along the DNA template.

Interactions between subsequent RNAPs that move along the same DNA segment have to be taken into consideration when more than one RNAP molecule initiates from the same promoter sequence of the DNA template. Pausing RNAP may block the advancement of trailing RNAP and thus induce “traffic jams” [13–15] that slow down elongation. On the other hand, the interaction may also be cooperative: Trailing RNAP can prevent backtracking, and even “push” the leading RNAP out of pause sites [16,17]. Thus elongation is enhanced.

This intriguing and seemingly paradoxical outcome of RNAP interactions has been studied intensely over the last few years using a variety of different approaches [18–25], mostly focusing on the role of specific microscopic features of the step cycle. To understand better the conditions under which jamming and pushing can arise from RNAP interactions we introduced a lattice gas model [26] that shows that indeed both phenomena arise if (in addition to pure steric hindrance on contact) RNAPs interact via a repulsive short-range interaction, while hard core repulsion alone can only produce jamming, see, e.g., [14,27–34] for ribosomes, RNAP, and other molecular motors.

Our model is a generalization of the asymmetric simple exclusion process (ASEP) [9,35], that incorporates both the two internal states in which RNAP appears (with or without PP_i bound to it) as in the pioneering work [14] and

*belitsky@ime.usp.br

†g.schuetz@fz-juelich.de

short-range interactions due to elastic deformation of the TEC upon collision [18,21]. Unlike in the works mentioned above, our starting point is a given probability distribution in parametric form from which we compute local transition rates for elongation and PP_i release such that this distribution is stationary, without requiring detailed balance since the transcription elongation takes place far from thermal equilibrium. This approach enables us to make *exact* analytical computations of transition rates in terms of interaction parameters and for collective stationary quantities.

It turns out [26] that any short-range repulsion leads to an enhanced microscopic forward translocation rate of an RNAP if a trailing RNAP has arrived to its left (i.e., from the backward direction), thus leading to *stochastic pushing*. However, this microscopic stochastic pushing that occurs on the level of individual jumps leads to an enhancement of the average velocity of the RNAP along the chain and thus to a *collective pushing* effect only if the repulsive interaction is sufficiently strong and the RNAP density is not too high. For higher densities, jamming due to microscopic blocking prevails. Indeed, the density dependence of collective pushing is quite intricate: Depending on the precise form of the microscopic interaction, it may exist only in a specific window of RNAP densities. For sufficiently strong repulsion the average elongation rate can exhibit two maxima at different RNAP densities. The model also predicts the dependence of the motor velocity and elongation rate on the concentration of NTP and the rate of PP_i release. We also note that the inverse average elongation rate is the mean time-headway between the crossing of the same DNA segment by two successive RNAP's [36,37]. Thus one obtains information on the experimentally relevant time-headway distribution.

The focus of the present paper is on the effect of interactions on the stationary fluctuations of the RNAP positions along the DNA track due to the inherently noisy dynamics of the translocation. First we discuss the space-headway distribution, i.e., the distribution of distances between subsequent RNAP that initiated at the same promoter sequence. The second result concerns the local density fluctuations that result not only from the randomness of the translocation, but also from the fluctuations in the total number density of RNAP along the DNA track.

In the next section we present the mathematical model of [26]. In Sec. III the new results on the headway distribution and the RNAP density fluctuations are derived and discussed. (Some mathematical details of the derivation are transferred to the Appendix.) Finally, in Sec. IV a summary of the results is given and some conclusions are drawn.

II. STOCHASTIC MODEL FOR INTERACTING RNAP

The typical size of a transcription bubble is around 15 bp [4,38] whereas the TEC covers a DNA segment of up to 35 bp. To highlight the effect of interactions we do not describe the various transformations of the TEC during each elongation step [7]. We simplify the complicated geometry of the TEC by representing it as a rod covering ℓ lattice sites, where ℓ is parameter of our model, and not differentiate between the TEC and the RNAP.

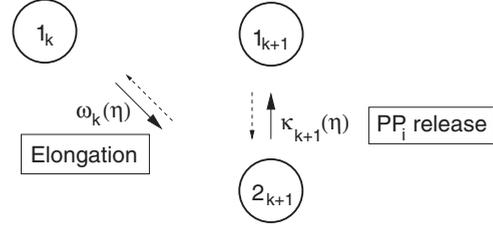


FIG. 1. Minimal reaction scheme of RNAP translocation. The RNAP i can move from base pair k to $k + 1$ provided it is in state 1_k (no pyrophosphate bound to it) with a configuration-dependent rate $\omega_i(\eta)$ through a series of processes involving NTP binding to the active site and NTP hydrolysis which results in a PP_i bound state in the transcription elongation complex (state 2_{k+1}). Only after PP_i release with a configuration-dependent rate $\kappa_i(\eta)$ (transition from state 2_{k+1} to state 1_{k+1}) the RNAP can perform the next translocation step. Reverse reactions (not considered in this work) are indicated with thin dashed arrows.

A. Stochastic dynamics

We follow [14] and base our model on the reduced description of Wang *et al.* [4] in which the rate-limiting step of the mechanochemical cycle is the PP_i release from the catalytic site [39]. Thus an RNAP appears in only two distinct polymerization states, namely without PP_i bound to it (state 1) or with PP_i (state 2). It is then convenient to characterize the state of a TEC mathematically not in terms of the length k of the RNA transcript attached to it, but to describe it in terms of the corresponding base pair so that $x = k$ marks the position of the RNAP on the template DNA. The RNAP moves forward along the DNA by one bp (a step length of $\delta = 0.34$ nm), i.e., from x to $x + 1$ only after PP_i release, i.e., only of the RNAP is in state 1_x . Without loss of generality we define x to be the lattice position of the left end of the rod in the random walk model.

In the presence of more than one RNAP, the symbol η denotes the full configuration of all RNAPs on the same DNA segment, i.e., the set of positions x_i and states α_i of the RNAPs, labeled successively by an integer $i = 1, 2, \dots, N$. We denote the rate at which an elongation step of RNAP i occurs by $\omega_i(\eta)$. The rate of PP_i release is denoted by $\kappa_i(\eta)$. The reverse processes occur with smaller rates [4] that we neglect. This minimal reaction scheme is sketched in Fig. 1 for a single RNAP.

In the rod picture, $x_i + \ell - 1$ is the lattice position of the “front” side of the TEC. The hard-core repulsion of RNAP implies that a forward move of an RNAP on site x_i by one site can occur only if the target site $x_i + \ell$ is not already covered. We say that two RNAP i and $i + 1$ are neighbors when the front of rod i and the left edge of rod $i + 1$ occupy neighboring lattice sites, i.e., when $x_{i+1} = x_i + \ell$. Thus the rates are in the form

$$\omega_i(\eta) = \omega^* \delta_{s_i, 1} (1 - \delta_{x_{i+1}, x_i + \ell}) (1 + d^{1*} \delta_{x_{i-1} + \ell, x_i} + d^{*1} \delta_{x_{i+1}, x_i + \ell + 1}), \quad (2.1)$$

$$\begin{aligned} \kappa_i(\eta) = & \kappa^* \delta_{s_i, 2} [1 + f^{1*} \delta_{x_{i-1} + \ell, x_i} + f^{*1} \delta_{x_i + \ell, x_{i+1}} \\ & + f^{1*1} \delta_{x_{i-1} + \ell, x_i} \delta_{x_i + \ell, x_{i+1}} + f^{10*} (1 - \delta_{x_{i-1} + \ell, x_i}) \\ & \times \delta_{x_{i-1} + \ell + 1, x_i} + f^{*01} (1 - \delta_{x_i + \ell, x_{i+1}}) \delta_{x_{i+1}, x_i + \ell + 1}], \end{aligned} \quad (2.2)$$

where in the setting of Wang *et al.* [4] the “bare” rates κ^* , ω^* when both neighboring sites $x_i - 1$ and $x_{i+\ell}$ are empty take the values

$$\kappa^* = 31.4s^{-1}, \quad \omega^* = [\text{NTP}](\mu M)^{-1}s^{-1}. \quad (2.3)$$

Here [NTP] is the NTP concentration which is parameter of our model. Under cellular conditions, RNAP transcription leads to downstream supercoiling that generates a load force of $6pN$ [40]. We ignore the effect of this load force as it only renormalizes the parameter ω^* . The phenomenological dimensionless parameters d^{1*} , d^{*1} and f^{1*} , f^{*1} , f^{1*1} , f^{10*} , f^{*01} describe the interaction between neighboring RNAP. The overall factor $(1 - \delta_{x_{i+1}, x_i+\ell})$ forbids jumps onto an occupied site.

The master equation for the probability $P_t(\eta)$ of finding the rods at time t in a configuration η thus reads

$$\frac{d}{dt}P_t(\eta) = \sum_{i=1}^N (\mathcal{L}_i^* P_t)(\eta), \quad (2.4)$$

with the adjoint \mathcal{L}_i^* of the single-particle Markov generator \mathcal{L}_i given by

$$(\mathcal{L}_i^* P_t)(\eta) = \omega_i(\eta_{il}^i)P_t(\eta_{il}^i) + \kappa_i(\eta_{rel}^i)P_t(\eta_{rel}^i) - [\omega_i(\eta) + \kappa_i(\eta)]P(\eta). \quad (2.5)$$

Here η_{il}^i is the configuration that leads to η before a translocation of RNAP i (i.e., with coordinate $x_i^{il} = x_i - 1$ and state $s_i^{il} = 3 - s_i$) and η_{rel}^i is the configuration η before PP_i release at RNAP i (i.e., with coordinate $x_i^{rel} = x_i$ and state $s_i^{rel} = 3 - s_i$).

The probability distribution $P_t(\eta)$ is defined to respect hardcore repulsion which means that configurations such that $x_{i+1} = x_i + \ell - 1$ are forbidden. Moreover, since TECs cannot overtake each other, only configurations satisfying the ordering condition $x_{i+1} \geq x_i + \ell$ have nonzero probability. We shall refer to such configurations as allowed configurations. An allowed configuration η is thus specified by a coordinate vector $\mathbf{x} = (x_1, \dots, x_N)$ with ordered integer coordinates and a state vector $\mathbf{s} = (s_1, \dots, s_N)$ with state variables $s_i \in \{1, 2\}$.

Since we are interested only in the elongation stage of transcription we take a lattice of L sites with periodic boundary conditions. The positions x_i of the rods are counted modulo L and labels i are counted modulo N . Therefore we denote by $\delta_{x,y}^M$ the Kronecker symbol with arguments x, y understood modulo M . We denote by N^α the fluctuating number of RNAPs in state $\alpha \in \{1, 2\}$ so that $N = N^1 + N^2$. We also define the excess $B(\mathbf{s}) = N^1 - N^2$.

B. Stationary distribution

For determining the stationary solution $\pi(\eta)$ of the many-particle master equation (2.4) one cannot impose detailed balance as on the timescale of translocation and PP_i release the system is not in thermal equilibrium. However, as shown in [26], a lattice divergence condition

$$\pi^{-1}(\eta)(\mathcal{L}_i^* \pi)(\eta) = \Phi_{i-1}(\eta) - \Phi_i(\eta) \quad \forall \eta \quad (2.6)$$

allows for determining local transition rates of the desired form (2.1) and (2.2) with the stationary Boltzmann

weights

$$\pi(\eta) = \exp \left[-\frac{1}{k_B T} (U + \lambda B) \right]. \quad (2.7)$$

Here U is a short-range interaction energy

$$U(\mathbf{x}) = J \sum_{i=1}^N \delta_{x_{i+1}, x_i+\ell}^L \quad (2.8)$$

analogous to the internal energy in mechanical systems. Positive J corresponds to repulsion. The chemical potential λ is a Lagrange multiplier for the fluctuations in the excess

$$B(\mathbf{s}) = \sum_{i=1}^N (3 - 2s_i) \quad (2.9)$$

due to the stochasticity of NTP hydrolysis and PP_i release.

We introduce

$$x = e^{\frac{2\lambda}{k_B T}}, \quad y = e^{\frac{J}{k_B T}}, \quad (2.10)$$

so that $x > 1$ corresponds to an excess of RNAP in state 1 and repulsive interaction corresponds to $y > 1$. The normalized stationary distribution for allowed configurations is then given by

$$\pi^*(\eta) = \frac{1}{Z} y^{-\sum_{i=1}^N \delta_{x_{i+1}, x_i+\ell}^L} x^{-\sum_{i=1}^N (3/2 - s_i)} \quad (2.11)$$

with the partition function

$$Z = \sum_{\eta} \pi(\eta). \quad (2.12)$$

With the form (2.11) of the stationary distribution the condition (2.6) determines the parameters that enter the interaction the parameters appearing in (2.1) and (2.2) as follows [26]:

$$x = \frac{\omega^*}{\kappa^*}, \quad (2.13)$$

$$y = \frac{1 + d^{1*}}{1 + d^{*1}}, \quad (2.14)$$

$$f^{1*} = d^{1*} \frac{x}{1+x} - \frac{1}{1+x}, \quad (2.15)$$

$$f^{*1} = d^{1*} \frac{1}{1+x} - \frac{x}{1+x}, \quad (2.16)$$

$$f^{1*1} = -d^{1*}, \quad (2.17)$$

$$f^{10*} = d^{*1} \frac{1}{1+x}, \quad (2.18)$$

$$f^{*01} = d^{*1} \frac{x}{1+x}. \quad (2.19)$$

Thus the bare PP_i release rate κ^* , which sets the timescale of the process, and the hopping rates (given by ω^* and the interaction parameters d^{1*} , d^{*1}) are free parameters of the model.

III. RESULTS AND DISCUSSION

Due to the stochastic dynamics all quantities of interest are fluctuating. We use the exact stationary distribution (2.11) to study how some characteristic fluctuation properties of the

interacting RNAP depend on the density of RNAP and on the interaction strength. We define ρ to be the number density of RNAPs transcribing on the same DNA template. The model allows for a number density in the range $0 < \rho < 1/\ell$. In all plots shown below we have taken $\ell = 5$.

A. Headway distribution

The fluctuations in the position of the RNAPs manifest themselves in fluctuations of their *headway*, i.e., the number of empty sites m_i between neighboring rods i and $i + 1$. The distribution of the headway between the front of a trailing rod i and the back of a leading rod $i + 1$ is defined by

$$P_h(r) = \frac{1}{\rho} \langle \delta_{x_{i+1}-x_i-\ell, r} \rangle. \quad (3.1)$$

This quantity is most conveniently computed using the headway representation of the process which arises from the fact that one may describe the RNAP model in terms of the headway distances which are number of vacant sites $m_i = x_{i+1} - (x_i + \ell) \bmod L$ between the left edge of rod $i + 1$ and the right edge of rod i and the total number of vacant sites $M = L - N$. Then the stochastic variables are $\zeta = (\mathbf{m}, \mathbf{s})$ with the headway distances $\mathbf{m} = (m_1, \dots, m_N)$.

With the indicator functions

$$\theta_i^p := \delta_{m_i, p} = \delta_{x_{i+1}, x_i + \ell + p}, \quad (3.2)$$

where the index i taken modulo N the transition rates (2.1) and (2.2) for extended interaction range become

$$\tilde{\omega}_i(\zeta) = \omega^* \delta_{s_i, 1} (1 - \theta_i^0) (1 + d^{1*} \theta_{i-1}^0 + d^{*1} \theta_i^1), \quad (3.3)$$

$$\begin{aligned} \tilde{\kappa}_i(\zeta) = & \kappa^* \delta_{s_i, 2} (1 + f^{1*} \theta_{i-1}^0 + f^{*1} \theta_i^0 + f^{1*1} \theta_{i-1}^0 \theta_i^0 \\ & + f^{10*} \theta_{i-1}^1 + f^{*01} \theta_i^1). \end{aligned} \quad (3.4)$$

In the mapping to the headway process the stationary average speed of an RNAP is given by the stationary expectation of the function $\tilde{\omega}_i(\zeta)$.

A translocation of RNAP i corresponds to the transition $(m_{i-1}, m_i) \rightarrow (m_{i-1} + 1, m_i - 1)$. The configurations that

lead to a given configuration ζ are $\zeta^{i-1, i}$ for translocation and ζ^i for PP_i release, defined by

$$\mathbf{s}_j^{i-1, i} = \mathbf{s}_j + (3 - 2s_j) \delta_{j, i}, \quad \mathbf{m}_j^{i-1, i} = m_j + \delta_{j, i} - \delta_{j, i-1}, \quad (3.5)$$

$$\mathbf{s}_j^i = \mathbf{s}_j + (3 - 2s_j) \delta_{j, i}, \quad \mathbf{m}_j^i = m_j. \quad (3.6)$$

This yields the master equation

$$\frac{d}{dt} P(\zeta, t) = \sum_{i=1}^N Q_i(\zeta, t), \quad (3.7)$$

with

$$\begin{aligned} Q_i(\zeta, t) = & \tilde{\omega}_i(\zeta^{i-1, i}) P(\zeta^{i-1, i}, t) - \tilde{\omega}_i(\zeta) P(\zeta, t) + \tilde{\kappa}_i(\zeta^i) \\ & \times P(\zeta^i, t) - \tilde{\kappa}_i(\zeta) P(\zeta, t). \end{aligned} \quad (3.8)$$

This is a generalized misanthrope process [41] where sites i can take two degrees of freedom s_i .

Since the sum of headways $\sum_{i=1}^N m_i = M$ is conserved, the stationary distribution of the headway process can be expressed in grandcanonical form with a fugacity \tilde{z} for the headways. In terms of the distance variables the stationary distribution (2.11) reads

$$\tilde{\pi}(\zeta) = \frac{1}{\tilde{Z}} \prod_{i=1}^N (x^{-3/2+s_i} y^{-\theta_i^0} \tilde{z}^{-m_i}). \quad (3.9)$$

The factorization indicates the absence of distance correlations. The partition function is given by

$$\tilde{Z} = \left[(x^{\frac{1}{2}} + x^{-\frac{1}{2}}) \frac{1 + (y-1)\tilde{z}}{1 - \tilde{z}} \right]^N, \quad (3.10)$$

which follows from the factorization of the distribution (3.9). The headway fugacity is given in terms of the RNAP density by

$$\tilde{z} = 1 - \frac{1 - (\ell - 1)\rho - \sqrt{[1 - (\ell - 1)\rho]^2 - 4\rho(1 - \ell\rho)(1 - y^{-1})}}{2(1 - \ell\rho)(1 - y^{-1})}. \quad (3.11)$$

Since $0 < \rho < 1/\ell$ one has $0 \leq \tilde{z} < 1$ so that the partition function (3.9) is well defined.

From the factorization property of (3.9) one obtains immediately the headway distribution

$$P_h(r) = \begin{cases} \frac{1 - \tilde{z}}{1 + (y-1)\tilde{z}} & \text{for } r = 0, \\ y P_h(0) \tilde{z}^r & \text{for } r \geq 1, \end{cases} \quad (3.12)$$

with \tilde{z} given in (3.11) as a function of the RNAP density ρ . One recognizes an ‘‘almost’’ geometric distribution which differs from a genuine geometric distribution only by a deviation at distance 0, i.e., for neighboring rods. For $y = 1$ one has $\tilde{z} = (1 - \ell\rho)/(1 - (\ell - 1)\rho)$ and the distribution is geometric. For repulsive interaction nearest-neighbor configurations

are suppressed. This is demonstrated in Fig. 2. The headway probability $P_h(r)$ is larger than in the noninteracting for any $r \geq 1$, while for neighboring rods ($r = 0$) it is smaller. In the attractive case one observes the opposite property.

The quantity

$$r^* = -(\ln \tilde{z})^{-1} \quad (3.13)$$

defines the localization length which characterizes the decay of the headway distribution. It depends strongly on the density, which is not surprising since, for short-range interactions, one expects the localization length r^* to be of similar magnitude as the average headway M/N . Indeed, in the noninteracting case one gets from (3.12) $\langle r \rangle = z/(1 - z)$ and from (3.11) one finds $\tilde{z} = 1 - \rho/[1 - (\ell - 1)\rho]$ so that

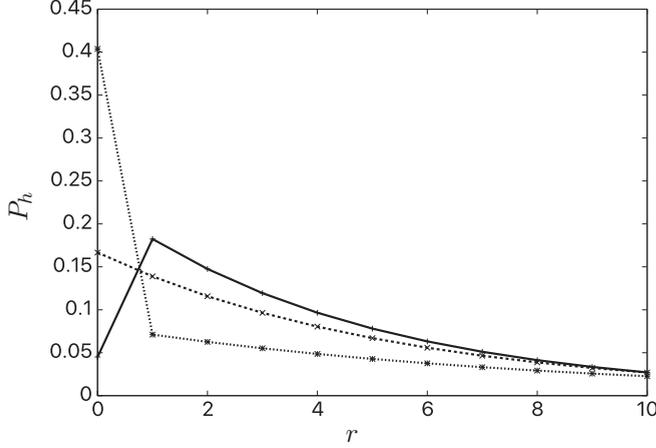


FIG. 2. RNAP headway distribution $P_h(r)$ for different interaction strengths y as a function of the integer lattice distance. Full curve: $y = 5.0$ (strong repulsion); Dashed curve: $y = 1.0001$ (only hard core); Dotted curve: $y = 0.2$ (attraction). The curves joining the data points are guides to the eye.

$\langle r \rangle = (1 - \ell\rho)/\rho = M/N$. For fixed ρ , there is a moderate dependence on the interaction strength in the attractive case. For repulsion, however, the localization length is not very sensitive to the interaction strength, see Fig. 3.

B. Density correlations

The headway distribution does not give immediate insight into the RNAP density correlations, i.e., into the probability $C(r)$ of having any two rods at a distance r . To study the stationary expectation of this fluctuating quantity we focus on the situation where the length of the rod is comparable to the interaction range, i.e., we consider $\ell = 1$.

Next we introduce the occupation number representation $\eta = (\eta_1, \eta_2, \dots, \eta_L)$ in which site k of the lattice is either empty ($\eta_k = 0$) or covered by a rod in state s_k in which case

one has $\eta_k = s_k$. We also the indicator function

$$n_k^\alpha := \delta_{\eta, \alpha} \quad (3.14)$$

and the variables

$$n_k := n_k^1 + n_k^2, \quad \sigma_k := n_k^1 - n_k^2. \quad (3.15)$$

Turning to a grandcanonical description the invariant measure (3.9) written in terms of the occupation variables σ_k, n_k becomes

$$\hat{\pi}(\eta) = \frac{1}{\hat{Z}} \prod_{k=1}^L (x^{-\frac{1}{2}\sigma_k} y^{-n_k n_{k+1}} z^{-n_k}). \quad (3.16)$$

Here the particle fugacity $z = \exp(\beta\mu)$ is a Lagrange multiplier with chemical potential μ that accounts for the fluctuating total number of RNAP along the DNA template.

One recognizes in (3.16) a generalized Ising measure that can be treated with textbook transfer matrix techniques, see the Appendix where we express the invariant measure (3.16) in terms of a three-dimensional transfer matrix \hat{T} . The partition function \hat{Z} and the density correlation functions

$$C^{\alpha\beta}(r) := \langle n_k^\alpha n_{k+r}^\beta \rangle - \langle n_k^\alpha \rangle \langle n_{k+r}^\beta \rangle \quad (3.17)$$

are the conveniently computed by diagonalizing \hat{T} .

With the matrices

$$\hat{n}^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \hat{n}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (3.18)$$

and $\hat{n} = \hat{n}^1 + \hat{n}^2$ one then obtains from the transfer matrix representation (A6) of the invariant measure

$$\rho = \langle n_k \rangle = \frac{1}{\hat{Z}} \text{Tr}(\hat{T}^{k-1} \hat{n} \hat{T}^{L_s - k + 1}) \quad (3.19)$$

as a function of the chemical potential μ . Because of translation invariance the pseudodensity does not depend on k . Diagonalizing the transfer matrix one gets for general joint expectations the expression

$$\langle n_{k_1}^{\eta_1} n_{k_2}^{\eta_2} \dots n_{k_p}^{\eta_p} \rangle_{L_s} = \prod_{i=1}^p |A_+^{\eta_i}|^2 \prod_{i=1}^{p-1} \left[1 + \frac{A_-^{\eta_i} A_-^{\eta_{i+1}}}{A_+^{\eta_i} A_+^{\eta_{i+1}}} \left(\frac{\chi_+}{\chi_-} \right)^{-(k_{i+1} - k_i)} \right] \frac{1 + \frac{A_-^{\eta_p} A_-^{\eta_1}}{A_+^{\eta_p} A_+^{\eta_1}} \left(\frac{\chi_+}{\chi_-} \right)^{-(L_s + k_1 - k_p)}}{1 + \left(\frac{\chi_+}{\chi_-} \right)^{-L_s}} \quad (3.20)$$

with the eigenvalues χ_\pm (A9) and eigenvector components $A_\pm^{\eta_i}$ (A10) of the transfer matrix \hat{T} (A5).

In the thermodynamic limit one has

$$\lim_{L \rightarrow \infty} \left(\frac{\chi_-}{\chi_+} \right)^L = 0 \quad (3.21)$$

since $\chi_- < \chi_+$ for any finite μ . For $k_p = o(L)$ this yields

$$\langle n_{k_1}^{\eta_1} n_{k_2}^{\eta_2} \dots n_{k_p}^{\eta_p} \rangle = \left[\prod_{i=1}^{p-1} (A_+^{\eta_i} A_+^{\eta_{i+1}} \chi_+^{k_{i+1} - k_i} + A_-^{\eta_i} A_-^{\eta_{i+1}} \chi_-^{k_{i+1} - k_i}) \right] A_+^{\eta_p} A_+^{\eta_1} \chi_+^{k_1 - k_p} \quad (3.22)$$

$$= \prod_{i=1}^p |A_+^{\eta_i}|^2 \prod_{i=1}^{p-1} \left[1 + \frac{A_-^{\eta_i} A_-^{\eta_{i+1}}}{A_+^{\eta_i} A_+^{\eta_{i+1}}} \left(\frac{\chi_+}{\chi_-} \right)^{-(k_{i+1} - k_i)} \right] \quad (3.23)$$

$$\langle n_1^{\eta_1} n_2^{\eta_2} \dots n_p^{\eta_p} \rangle = \left(\prod_{i=1}^{p-1} \hat{T}_{\eta_i \eta_{i+1}} \right) A_+^{\eta_p} A_+^{\eta_1} \chi_+^{1-p}. \quad (3.24)$$

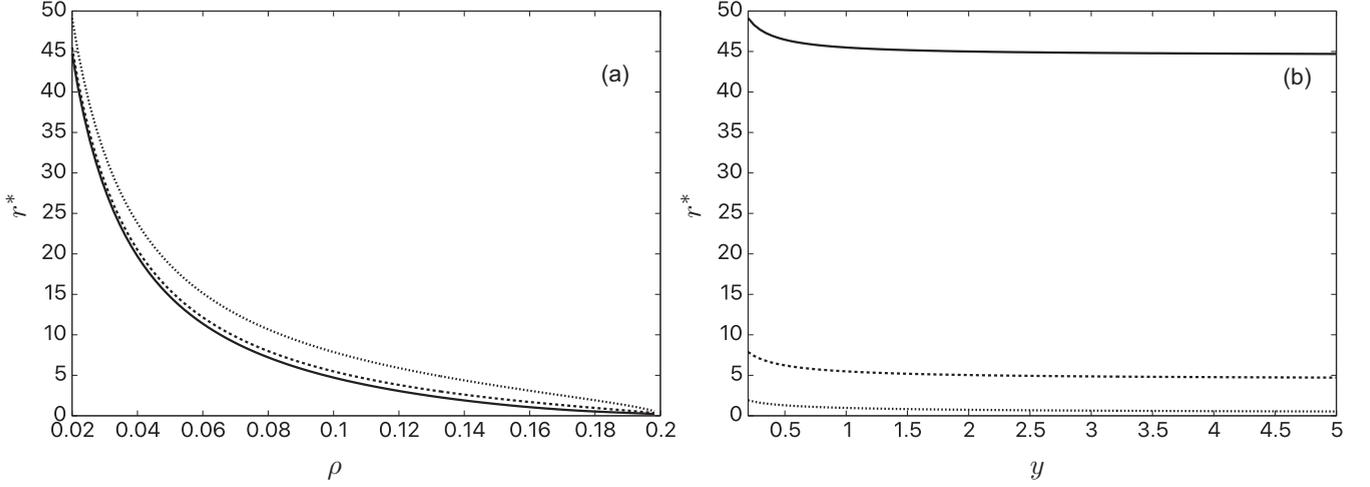


FIG. 3. RNAP localization length $r^*(\rho, y)$ as function of RNAP density for different interaction strengths $y = 5.0, 1.0, 0.2$ [(a), curves from bottom to top] and as function of the interaction strength y for different densities $\rho = 0.02, 0.1, 0.18$ [(b), curves from bottom to top].

More specifically, from the eigenvectors (A10) and eigenvalues (A9) one finds

$$\rho^1 = \langle n_k^1 \rangle = \frac{1}{1+x} \rho, \quad (3.25)$$

$$\rho^2 = \langle n_k^2 \rangle = \frac{x}{1+x} \rho, \quad (3.26)$$

and

$$\rho = \frac{1}{2} \left[1 + \frac{y^{-1}q - 1}{\sqrt{(y^{-1}q - 1)^2 + 4q}} \right], \quad (3.27)$$

which expresses ρ in terms of the chemical potential μ and the model parameters x, y . Conversely one has

$$q = y \left[1 + \frac{y(2\rho - 1)^2}{2\rho(1 - \rho)} + \frac{y(2\rho - 1)}{2\rho(1 - \rho)} \sqrt{4\rho(1 - \rho)y^{-1} + (2\rho - 1)^2} \right]. \quad (3.28)$$

For the correlation function the general expression (3.22) yields

$$C^{\alpha\beta}(r) = \frac{\rho^\alpha}{\rho} \frac{\rho^\beta}{\rho} \rho(1 - \rho) \frac{\chi_+^{-r}}{\chi_-^{-r}} \quad (3.29)$$

in terms of the eigenvalues (A9) of the transfer matrix. To express the correlation length in terms of the RNAP density we use (3.27) to obtain

$$y^{-1}q - 1 = \frac{y(2\rho - 1)}{2\rho(1 - \rho)} [2\rho - 1 + \sqrt{1 + 4\rho(1 - \rho)(y^{-1} - 1)}]. \quad (3.30)$$

On the other hand, (3.28) gives

$$\sqrt{(y^{-1}q - 1)^2 + 4q} = \frac{y}{2\rho(1 - \rho)} [2\rho - 1 + \sqrt{1 + 4\rho(1 - \rho)(y^{-1} - 1)}]. \quad (3.31)$$

Thus

$$\chi_{\pm} = 1 + \frac{y}{4\rho(1 - \rho)} (2\rho - 1 \pm 1) [2\rho - 1 + \sqrt{1 + 4\rho(1 - \rho)(y^{-1} - 1)}] \quad (3.32)$$

and we arrive at

$$e^{1/\xi} = 1 + \frac{\sqrt{1 + 4\rho(1 - \rho)(y^{-1} - 1)} + 1}{2\rho(1 - \rho)(y^{-1} - 1)}. \quad (3.33)$$

Notice that the eigenvalue χ_- changes sign at $y = 1$ for all ρ (Fig. 4).

For repulsive interaction ($y > 1$) one has $\chi_- < 0$ and we can define a correlation length ξ by

$$e^{1/\xi} = \left| \frac{\chi_+}{\chi_-} \right| \quad (3.34)$$

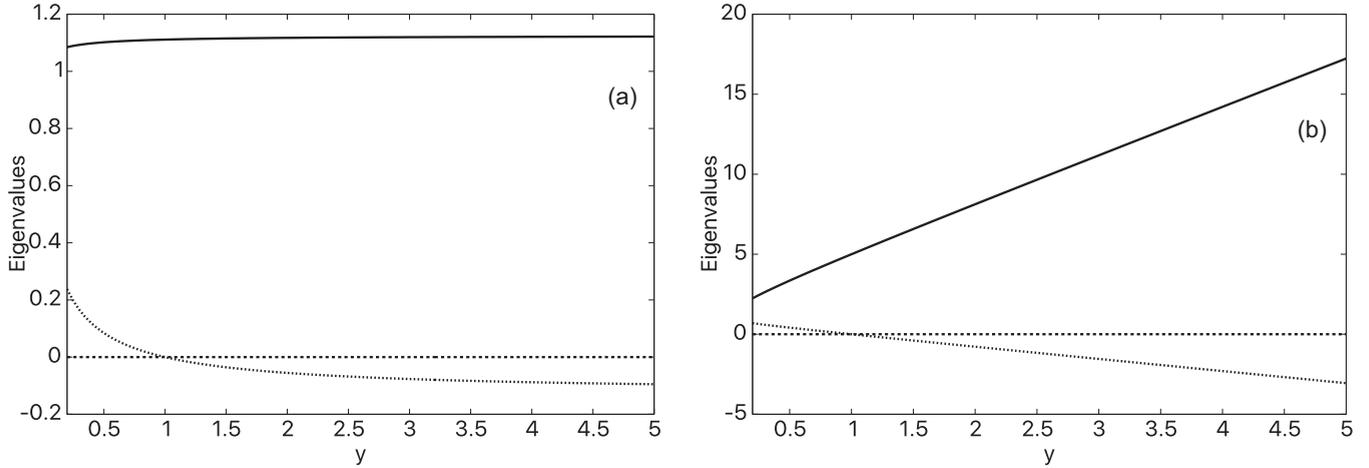


FIG. 4. Eigenvalues χ_{\pm} and $\chi_0 = 0$ as function of the interaction strength y for two RNAP densities (a) $\rho = 0.1$ and (b) $\rho = 0.8$. The solid line shows χ_+ . The eigenvalue χ_- changes sign at $y = 1$.

so that

$$C^{\alpha\beta}(r) = \frac{\rho^\alpha \rho^\beta}{\rho} \rho(1-\rho)(-1)^r e^{-r/\xi}. \quad (3.35)$$

One sees that the RNAP density correlations are staggered and decay exponentially fast, but with a correlation length that is much smaller than the localization length r^* (3.13) appearing on the headway distribution, see Fig. 5. One notices a significant dependence of the correlation length ξ on the density (except in the noninteracting case where the correlation length vanishes for all densities) and also a strong dependence on the interaction parameter for not too strong repulsion. As the repulsion gets stronger, the dependence of the correlation length flattens out. For attractive interaction ($y < 1$) one has $\chi_- > 0$ and therefore

$$C^{\alpha\beta}(r) = \frac{\rho^\alpha \rho^\beta}{\rho} \rho(1-\rho)e^{-r/\xi} \quad (3.36)$$

with the same decay of the amplitude as in the repulsive case but no staggering. The correlation length depends strongly on the interaction parameter for any $y < 1$.

IV. CONCLUSION

During transcription elongation multiple RNAPs may move one after another along the same DNA segment. In this work we study the stationary distribution of RNAP locations in terms of the headway distribution and the RNAP density correlation function, using the model developed in [26]. This model incorporates and explains the interplay of RNAP pushing and jamming as a function of the interaction strength and RNAP density. The model is based on a reduced picture of the mechanochemical cycle and incorporates a nearest-neighbor interaction. However, the approach is mathematically robust in the sense that it can be extended with the techniques of [26,41] to allow for a more detailed biological description of the mechanochemical cycle of the RNAP during

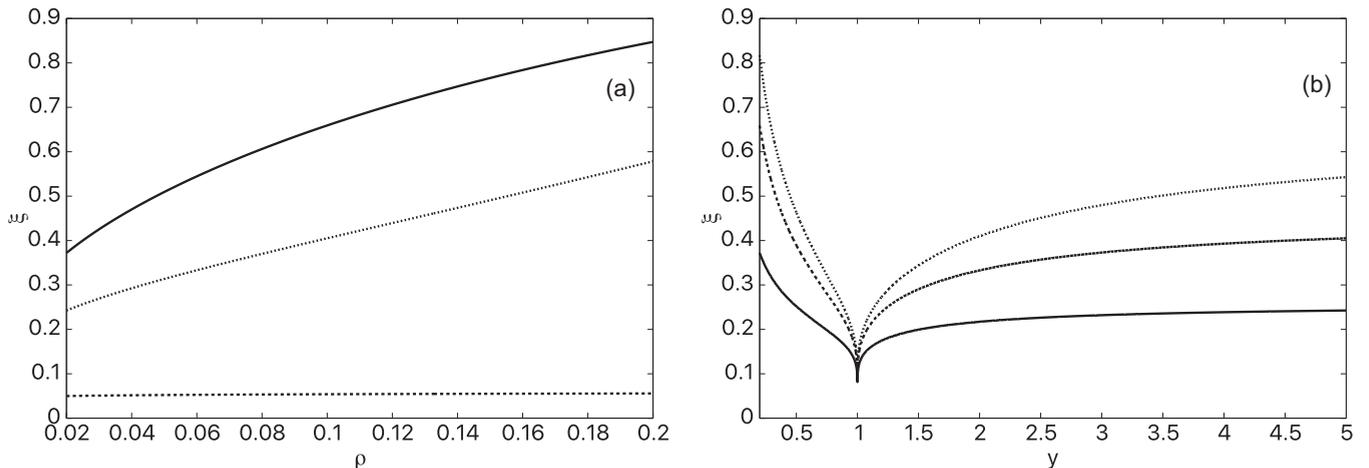


FIG. 5. RNAP density correlation length $\xi(\rho, y)$ as function of RNAP density (a) for different interaction strengths $y = 5.0$ (upper dashed curve), $y = 1.0$ (lower dashed curve), $y = 0.2$ (solid curve), and as function of the interaction strength y (b) for different densities $\rho = 0.02$ (solid curve at the bottom), $\rho = 0.1$ (upper dashed curve), $\rho = 0.18$ (lower dashed curve).

elongation and also for incorporating more general short-range interactions.

The model predicts (i) headways are uncorrelated and (ii) that the headway between successive RNAP is distributed geometrically beyond the interaction range which here was chosen, for simplicity, to extend only over one base pair. Intriguingly, this means that beyond the interaction range the distance to the next RNAP exhibits the same randomness as if no interactions were present. Conversely, the distribution at short distances may be a tool to probe interactions' strengths. This is important since one would expect an interaction range R of the order of the size of the TEC. The modeling approach chosen here can be extended to cover this scenario. For an invariant measure of the form (2.11), but with an extended interaction range R , one finds a geometric distribution for headways of size $r > R$, but not for headways $r \leq R$. The corrections for $r \leq R$ as a function of the generalized fugacity \tilde{z} follow directly from the interaction constants. The density ρ as a function of \tilde{z} is a polynomial equation of degree $R + 1$. Thus the interaction constants can be inferred by measuring the headway distribution for any density ρ . We note that, unlike in some related lattice gas models for vehicular traffic [42], the headway distribution exhibits (at most) a single peak, thus there is no phase coexistence between a congested region and a free-flow region in the present model for RNAP traffic.

A second quantity of interest turns out to be the density correlation function between RNAP. Interestingly, repulsive interaction leads to staggered density correlations, while an attractive interaction does not exhibit this phenomenon. The correlation length is very sensitive to the interaction strength, unless the interaction is strongly repulsive. Thus also the correlation length is a means to probe the nature of the interactions. For an interaction range $R > 1$ one expects a decay of correlations by a sum of exponentials with different correlation lengths.

Experimentally, one can regulate the number of RNAP initiating from the same promoter and hence the RNAP density by using an excess of DNA [19]. The interaction strength and range may be probed by applying an external torque to the RNAP [43]. It would also be interesting to derive the full experimentally important time-headway distribution between RNAP by generalizing the approach of [36,37] to the present case of additional short-range interactions in the particle hopping rate.

ACKNOWLEDGMENTS

G.M.S. thanks the Instituto de Matemática e Estatística of the University of São Paulo for kind hospitality. This work was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES),- Finance Code 001, by Grants No. 2017/20696-0 and No. 2017/10555-0 of the São Paulo Research Foundation (FAPESP), and by Grant No. 309239/2017-6 of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

APPENDIX: TRANSFER MATRIX FOR OCCUPATION NUMBER REPRESENTATION

The invariant measure (3.16) can be written as

$$\hat{\pi}(\boldsymbol{\eta}) = \frac{1}{\tilde{Z}} \prod_k T_{k,k+1} \quad (\text{A1})$$

with

$$T_{k,k+1}(\boldsymbol{\eta}) = y^{-n_k n_{k+1}} x^{-\frac{1}{4}(\sigma_k + \sigma_{k+1})} z^{-\frac{1}{2}(n_k + n_{k+1})}. \quad (\text{A2})$$

Introducing the three-dimensional canonical basis vectors of \mathbb{C}^3

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (\text{A3})$$

with the scalar product $\langle \alpha | \beta \rangle = \delta_{\alpha,\beta}$ one finds

$$T_{k,k+1}(\boldsymbol{\eta}) = \langle \eta_k | \hat{T} | \eta_{k+1} \rangle \quad (\text{A4})$$

with the 3×3 transfer matrix

$$\hat{T} = \begin{pmatrix} 1 & x^{-\frac{1}{4}} z^{-\frac{1}{2}} & x^{\frac{1}{4}} z^{-\frac{1}{2}} \\ x^{-\frac{1}{4}} z^{-\frac{1}{2}} & x^{-\frac{1}{2}} y^{-1} z^{-1} & y^{-1} z^{-1} \\ x^{\frac{1}{4}} z^{-\frac{1}{2}} & y^{-1} z^{-1} & x^{\frac{1}{2}} y^{-1} z^{-1} \end{pmatrix}. \quad (\text{A5})$$

It follows that

$$\hat{\pi}(\boldsymbol{\eta}) = \frac{1}{\tilde{Z}} \langle \eta_1 | \hat{T} | \eta_2 \rangle \langle \eta_2 | \hat{T} | \eta_3 \rangle \dots \langle \eta_{L'-1} | \hat{T} | \eta_{L'} \rangle \times \langle \eta_{L'} | \hat{T} | \eta_1 \rangle \quad (\text{A6})$$

and

$$\hat{Z} = \text{Tr} \hat{T}^{L'}. \quad (\text{A7})$$

We introduce the parameter

$$q = z^{-1}(x^{\frac{1}{2}} + x^{-\frac{1}{2}}). \quad (\text{A8})$$

The eigenvalues of \hat{T} are $\chi_0 = 0$ and

$$\chi_{\pm} = \frac{1}{2}(1 + qy^{-1}) \pm \frac{1}{2}\sqrt{(1 - qy^{-1})^2 + 4q}. \quad (\text{A9})$$

All eigenvalues are real in the physical parameter range and one has $\chi_+ > 1$ (with equality only if $\mu = \infty$). For $K \neq 0$ there is no degeneracy. For $K = 0$ one has $\chi_+ = 1 + 2e^{-\mu} \cosh \lambda$ and $\chi_- = \chi_0 = 0$.

The eigenvectors $|0, \pm\rangle$ with components $A_{0,\pm}^{\eta}$ for $\eta \in \{0, 1, 2\}$ are

$$|0\rangle = \frac{z^{-\frac{1}{2}}}{\sqrt{\kappa_0}} \begin{pmatrix} 0 \\ x^{\frac{1}{4}} \\ -x^{-\frac{1}{4}} \end{pmatrix}, \quad |\pm\rangle = \frac{1}{\sqrt{\kappa_{\pm}}} \begin{pmatrix} q \\ x^{-\frac{1}{4}} z^{-\frac{1}{2}} (\chi_{\pm} - 1) \\ x^{\frac{1}{4}} z^{-\frac{1}{2}} (\chi_{\pm} - 1) \end{pmatrix} \quad (\text{A10})$$

with the normalization factors

$$\kappa_0 = q, \quad \kappa_{\pm} = q[q + (\chi_{\pm} - 1)^2]. \quad (\text{A11})$$

- [1] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. K. Roberts, and P. Walter, *Essential Cell Biology*, 4th ed., (Garland Science, New York, 2013).
- [2] L. Bai, T. J. Santangelo, and M. D. Wang, Single-Molecule Analysis of RNA Polymerase Transcription, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 343 (2006).
- [3] C. Ma, X. Yang, and P. J. Lewis, Bacterial transcription as a target for antibacterial drug development, *Microbiol. Mol. Biol. Rev.* **80**, 139 (2016).
- [4] H. Y. Wang, T. Elston, A. Mogilner, and G. Oster, Force Generation in RNA Polymerase, *Biophys. J.* **74**, 1186 (1998).
- [5] L.-T. Da, Chao E., B. Duan, C. Zhang, X. Zhou, and J. Yu, A jump-from-cavity pyrophosphate ion release assisted by a key lysine residue in T7 RNA polymerase transcription elongation, *PLoS Comput. Biol.* **11**, e1004624 (2015).
- [6] Y. X. Mejia, E. Nudler, and C. Bustamante, Trigger loop folding determines transcription rate of *Escherichia coli*'s RNA polymerase, *Proc. Natl. Acad. Sci. USA* **112**, 743 (2015).
- [7] L. Zhang, F. Pardo-Avila, I. C. Unarta, P. P.-H. Cheung, G. Wang, D. Wang, and X. Huang, Elucidation of the dynamics of transcription elongation by RNA polymerase II using kinetic network models, *Acc. Chem. Res.* **49**, 687 (2016).
- [8] M. Guthold, X. Zhu, C. Rivetti, G. Yang, N. H. Thomson, S. Kasas, H. G. Hansma, B. Smith, P. K. Hansma, and C. Bustamante, Direct observation of one-dimensional diffusion and transcription by *Escherichia coli* RNA polymerase, *Biophys. J.* **77**, 2284 (1999).
- [9] A. Schadschneider, D. Chowdhury, and K. Nishinari, *Stochastic Transport in Complex Systems* (Elsevier, Amsterdam, 2010).
- [10] D. Ó Maoiléidigh, V. R. Tadigotla, E. Nudler, and A. E. Ruckenstein, A unified model of transcription elongation: what have we learned from single-molecule experiments? *Biophys. J.* **100**, 1157 (2011).
- [11] L. Bai, A. Shundrovsky, and M. D. Wang, Sequence-dependent kinetic model for transcription elongation by RNA polymerase, *J. Mol. Biol.* **344**, 335 (2004).
- [12] L. Bai, R. M. Fulbright, and M. D. Wang, Mechanochemical Kinetics of Transcription Elongation, *Phys. Rev. Lett.* **98**, 068103 (2007).
- [13] S. Klumpp and T. Hwa, Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination, *Proc. Natl. Acad. Sci. USA* **105**, 18159 (2008).
- [14] T. Tripathi and D. Chowdhury, Interacting RNA polymerase motors on a DNA track: Effects of traffic congestion and intrinsic noise on RNA synthesis, *Phys. Rev. E* **77**, 011921 (2008).
- [15] T. Tripathi, G. M. Schütz, and D. Chowdhury, RNA polymerase motors: dwell time distribution, velocity and dynamical phases, *J. Stat. Mech.* (2009) P08018.
- [16] V. Epshtein and E. Nudler, Cooperation between RNA polymerase molecules in transcription elongation, *Science* **300**, 801 (2003).
- [17] V. Epshtein, F. Toulme, A. Rachid Rahmouni, S. Borukhov, and E. Nudler, Transcription through the roadblocks: the role of RNA polymerase cooperation, *EMBO J.* **22**, 4719 (2003).
- [18] H. Saeki and J. Q. Svejstrup, Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes, *Mol. Cell.* **35**, 191 (2009).
- [19] J. Jin, L. Bai, D. S. Johnson, R. M. Fulbright, M. L. Kireeva, M. Kashlev, and M. D. Wang, Synergistic action of RNA polymerases in overcoming the nucleosomal barrier, *Nat. Struct. Mol. Biol.* **17**, 745 (2010).
- [20] M. Ehrenberg, P. P. Dennis, and H. Bremer, Maximum *rrn* promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase, *Biochimie* **92**, 12 (2010).
- [21] E. A. Galburt, J. M. R. Parrondo, and S. W. Grill, RNA polymerase pushing, *Biophys. Chem.* **157**, 43 (2011).
- [22] S. Klumpp, Pausing and backtracking in transcription under dense traffic conditions, *J. Stat. Phys.* **142**, 1252 (2011).
- [23] P. R. Costa, M. L. Acencio, and N. Lemke, Cooperative RNA polymerase molecules behavior on a stochastic sequence-dependent model for transcription elongation, *PLoS ONE* **8**, e57328 (2013).
- [24] H. Teimouri, A. B. Kolomeisky, and K. Mehrabiani, Theoretical analysis of dynamic processes for interacting molecular motors, *J. Phys. A* **48**, 065001 (2015).
- [25] T. Heberling, L. Davis, J. Gedeon, C. Morgan, and T. Gedeon, A Mechanistic Model for Cooperative Behavior of Co-transcribing RNA Polymerases, *PLoS Comput. Biol.* **12**, e1005069 (2016).
- [26] V. Belitsky and G. M. Schütz, RNA polymerase interactions and elongation rate, *J. Theor. Biol.* **462**, 370 (2019).
- [27] L. B. Shaw, R. K. P. Zia, and K. H. Lee, Totally asymmetric exclusion process with extended objects: A model for protein synthesis, *Phys. Rev. E* **68**, 021910 (2003).
- [28] S. Klumpp and R. Lipowsky, Traffic of molecular motors through tube-like compartments, *J. Stat. Phys.* **113**, 233 (2003).
- [29] T. Chou and G. Lakatos, Clustered Bottlenecks in mRNA Translation and Protein Synthesis, *Phys. Rev. Lett.* **93**, 198101 (2004).
- [30] E. Frey, A. Parmeggiani, and T. Franosch, Collective Phenomena in Intracellular Processes, *Genome Inform.* **15**, 46 (2004).
- [31] J. J. Dong, B. Schmittmann, and R. K. P. Zia, Towards a model for protein production rates, *J. Stat. Phys.* **128**, 21 (2007).
- [32] P. Greulich, A. Garai, K. Nishinari, A. Schadschneider, and D. Chowdhury, Intracellular transport by single-headed kinesin KIF1A: Effects of single-motor mechanochemistry and steric interactions, *Phys. Rev. E* **75**, 041905 (2007).
- [33] A. B. Kolomeisky and M. E. Fisher, Molecular motors: a theorist's perspective, *Annu. Rev. Phys. Chem.* **58**, 675 (2007).
- [34] D. Chowdhury, Stochastic mechano-chemical kinetics of molecular motors: A multidisciplinary enterprise from a physicist's perspective, *Phys. Rep.* **529**, 1 (2013).
- [35] G. M. Schütz, *Exactly Solvable Models for Many-Body Systems far from Equilibrium*, in *Phase Transitions and Critical Phenomena*, Vol. 19, edited by C. Domb and J. Lebowitz, (Academic, London, 2001).
- [36] K. Ghosh, A. Majumdar, and D. Chowdhury, Distribution of time-headways in a particle-hopping model of vehicular traffic, *Phys. Rev. E* **58**, 4012 (1998).
- [37] P. Hrabák and M. Krbálek, Time-headway distribution for totally asymmetric exclusion process with various updates, *Phys. Lett. A* **380**, 1003 (2016).

- [38] K. Geszvain and R. Landick, *The Structure of Bacterial RNA Polymerase in The Bacterial Chromosome*, edited by P. Higgins, (ASM, Washington, 2005), pp. 283–296.
- [39] D. A. Erie, T. D. Yager, and P. Von Hippel, The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 379 (1992).
- [40] H. Yin, M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles, Transcription against an applied force, *Science* **270**, 1653 (1995).
- [41] C. Cocozza-Thivent, Processus des misanthropes, *Z. Wahrsch. Verw. Gebiete* **70**, 509 (1985).
- [42] D. Chowdhury, A. Majumdar, K. Ghosh, S. Sinha, and R. B. Stinchcombe, Particle-hopping models of vehicular traffic: Distributions of distance headways and distance between jams, *Physica A (Amsterdam)* **246**, 471 (1997).
- [43] T. Tripathi, P. Prakash, and D. Chowdhury, RNA polymerase motor on DNA track: effects of interactions, external force and torque. [arXiv:0812.4692](https://arxiv.org/abs/0812.4692).