# Biologically motivated asymmetric exclusion process: Interplay of congestion in RNA polymerase traffic and slippage of nascent transcript

Soumendu Ghosh,[1,*] Annwesha Dutta,[1,*] Shubhadeep Patra,[2,†] Jun Sato,[3] Katsuhiro Nishinari,[3] and Debashish Chowdhury[1,‡]

[1]*Department of Physics, Indian Institute of Technology, Kanpur 208016, India*
[2]*ISERC, Visva-Bharati, Santiniketan 731235, India*
[3]*Research Center for Advanced Science and Technology, The University of Tokyo, Komaba 4-6-1, Meguro-ku, Tokyo 153-8904, Japan*

We develop a theoretical framework, based on an exclusion process, that is motivated by a biological phenomenon called transcript slippage (TS). In this model a discrete lattice represents a DNA strand while each of the particles that hop on it unidirectionally, from site to site, represents a RNA polymerase (RNAP). While walking like a molecular motor along a DNA track in a step-by-step manner, a RNAP simultaneously synthesizes an RNA chain; in each forward step it elongates the nascent RNA molecule by one unit, using the DNA track also as the template. At some special "slippery" position on the DNA, which we represent as a defect on the lattice, a RNAP can lose its grip on the nascent RNA and the latter's consequent slippage results in a final product that is either longer or shorter than the corresponding DNA template. We develop an exclusion model for RNAP traffic where the kinetics of the system at the defect site captures key features of TS events. We demonstrate the interplay of the crowding of RNAPs and TS. A RNAP has to wait at the defect site for a longer period in more congested RNAP traffic, thereby increasing the likelihood of its suffering a larger number of TS events. The qualitative trends of some of our results for a simple special case of our model are consistent with experimental observations. The general theoretical framework presented here will be useful for guiding future experimental queries and for analysis of the experimental data with more detailed versions of the same model.

## I. INTRODUCTION

Each single DNA strand is a heteropolymer whose monomeric subunits are called nucleotides. By convention, the letters "A," "T," "C," and "G," represent the four nucleotide bases of a DNA. The specific sequence in which these four letters appear in a DNA strand is a chemically encoded genetic message (genetic information). Transcription of the genetic message is carried out by a molecular machine called RNA polymerase (RNAP) [1]. A RNAP synthesizes a RNA molecule (the transcript) whose sequence of monomeric subunits is complementary to that of a specific segment of its DNA that encodes the corresponding genetic message. The four letters of the alphabet that store genetic messages in the RNA transcript are "A," "U," "C," and "G." Each RNAP can also be regarded as a molecular motor [2–4] for which the DNA template also serves as the track for its unidirectional, albeit noisy, movement during transcription [1]. In each step forward, by one unit, along its DNA track, the RNAP elongates the nascent RNA transcript by one unit where the unit is measured in terms of a nucleotide.

Experiments revealed the existence of specific stretches of DNA sequence where the nascent RNA may slip, backward or forward, with respect to the RNAP although the RNAP

motor itself does not slip simultaneously on its DNA track [5]. In fact, at any given slippage-prone site, multiple successive events of backward and/or forward slippage may occur before the correct transcription can resume. Thus, transcript slippage (TS) results in length heterogeneity of the final products of transcription because of the incorporation of more or fewer nucleotides, respectively, as compared to the length of the transcript encoded in the DNA. While this phenomenon has received much attention over the past few decades [5,6], the detailed mechanism of TS, its causes and consequences are still unclear. In this paper we focus on the consequences, rather than the causes, of TS.

Often the same segment of DNA (here loosely defined as the "gene") is simultaneously transcribed by several RNAPs, each synthesizing a distinct copy of the same RNA. The collective movement of multiple RNAPs simultaneously on a DNA track resembles, at least superficially, vehicular traffic on highways [7,8]. Wide varieties of collective trafficlike phenomena in nonliving as well as in living systems have been modeled by various appropriate extensions of the totally asymmetric simple exclusion process (TASEP) [7–11]. In the past, RNAP traffic has been modeled theoretically [12–18] by extending the TASEP [7,19–21]. A RNAP is expected to dwell longer at the slippery site in congested traffic because of the hindrance caused by the leading vehicle. The longer a RNAP dwells at the slippery site, the larger the number of TS events it is likely to suffer. Thus, traffic congestion can influence the extent of TS. Here we develop a TASEP-based model to investigate the interplay of TS and RNAP traffic.

---

*These authors contributed equally to this work.
†Undergraduate Visitor at IIT Kanpur.
‡Corresponding author: debch@iitk.ac.in

The TASEP can never be in thermodynamic equilibrium, but can attain nonequilibrium steady states (NESS) [19,22,23]. One of the key properties of the NESS is the nonvanishing particle flux which is defined as the number of particles passing through a site per unit time. The effects of different types of defects and inhomogeneities on the flux and the density profile of the particles have been investigated extensively over the last three decades [24–38]. We treat the "slippage-prone site," where TS occurs, as a special type of "defect" in a TASEP-based model of RNAP traffic.

In traffic engineering it is essential to first characterize the driving behaviors of individual drivers before embarking on a study of vehicular traffic on highways. In the same spirit, here we study the statistical characteristic of a single RNAP that undergoes TS at a specific location on the DNA track, before studying of RNAP traffic on the same track. Thus, in this paper, we consider two different situations. In the first, a single RNAP is assumed to be moving alone on the DNA track, whereas in the second, many RNAPs move simultaneously in the same direction on a single DNA track. More specifically, we study three aspects of TS: (a) how is the rate of transcription by a single RNAP affected by TS? (b) how is the error due to TS affected by the traffic congestion during the collective movement of RNAPs? and (c) how, in turn, is the collective trafficlike movement of RNAP affected by TS statistics?

To study the effect of a transcript slippage on the movement of each individual RNAP on the DNA track and, equivalently, the rate of transcription, we use the technique of calculating first passage time (FPT) [39–41]. With this objective, we first construct a stochastic kinetic theory that incorporates the effect of arbitrary numbers of backward and forward slippage. Then, as a concrete example, we consider a special case of the model that can be treated analytically without much mathematical difficulty. For this concrete case, we compute the time taken by the RNAP motor to traverse a slippery site *for the first time*. This time is intrinsically stochastic and is termed here as the first-passage time. We interpret the results physically to explain how the movement of a RNAP on the DNA template (track) is affected by TS.

To study the interplay of RNAP traffic and TS, using mean-field approximation (MFA), we again compute the mean time needed to traverse a slippage site in a traffic of RNAPs on the same DNA track. We also carry out Monte Carlo simulations of the model and compare the Monte Carlo simulation data with the corresponding mean-field theoretic predictions to test the level of accuracy of the MFA. Finally, we also compare the theoretically predicted probability distribution of the longer and shorter transcripts with the experimental data [42,43] obtained through advanced sequencing technologies [44].

The paper is organized in the following manner. In Sec. II, we begin by sketching a brief introduction to the phenomenon of TS, followed by the description of our stochastic kinetic model of TS. In Sec. III, we study the effect of TS on a single RNAP. More specifically, we derive an exact analytical expression for the mean time taken by a single RNAP, in the absence of steric hindrance from any other RNAP, to traverse the site where TS is likely to occur. In Sec. IV, we investigate the effect of RNAP traffic congestion on TS. Our results establish that, because of traffic congestion, on average, each

RNAP dwells for a longer time at the defect site and hence suffers a larger number of TS events. In Sec. V, we analyze the effect of TS on RNAP traffic flow; the average density profile exhibits features that are typical characteristics of the TASEP with point defects. Finally, in Sec. VI, we present a summary of the results and draw conclusions.

## II. MODEL AND BIOLOGICAL MOTIVATION

We begin this section with a brief overview of the TS process, as depicted schematically in Fig. 1, in Sec. II A. Then, motivated by this biological phenomenon, in Sec. II B we develop our theoretical model. The distinct kinetic states are displayed, and the interstate transitions are indicated, in Fig. 2.
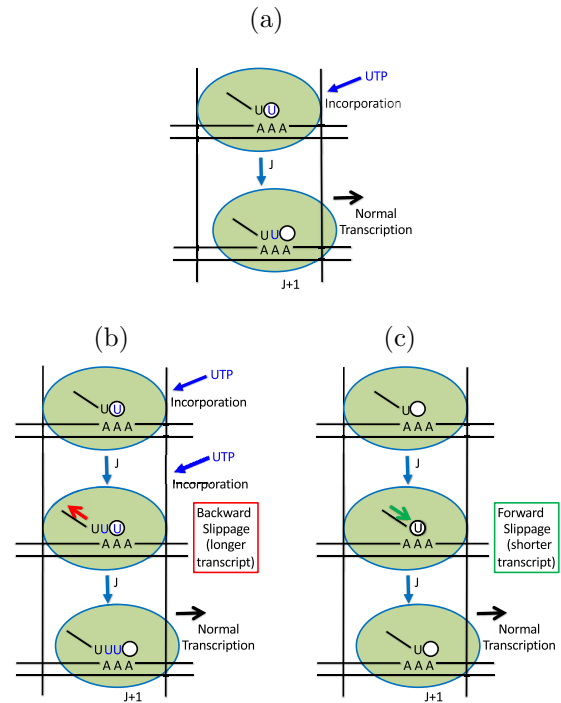


FIG. 1. A pictorial depiction of (a) normal transcription, without slippage, by an RNAP, (b) transcription with backward slippage of nascent RNA, and (c) transcription with forward slippage of nascent RNA. The white circle represents the active site of an RNAP and the blue letter "U" depicts the incorporation of a nucleotide in the active site. The array of nucleotides "A" represents slippery sequence in the DNA strand and the slanted black solid line represents nascent RNA. Backward and forward slippages of the nascent RNA transcript are indicated by red and green arrows, respectively. In (a), after incorporation of a nucleotide "U," RNAP can move one step forward with respect to the DNA template. In (b), after incorporation of a nucleotide "U," nascent RNA slips backward with respect to RNAP as well as the DNA template, by keeping RNAP fixed in its position. This results in the addition of an extra nucleotide "U" to the transcript. In (c), before the incorporation of a "U," nascent RNA can slip forward with respect to RNAP as well as the DNA template, by keeping RNAP fixed in its position, resulting in a shortening of the transcript by one nucleotide.
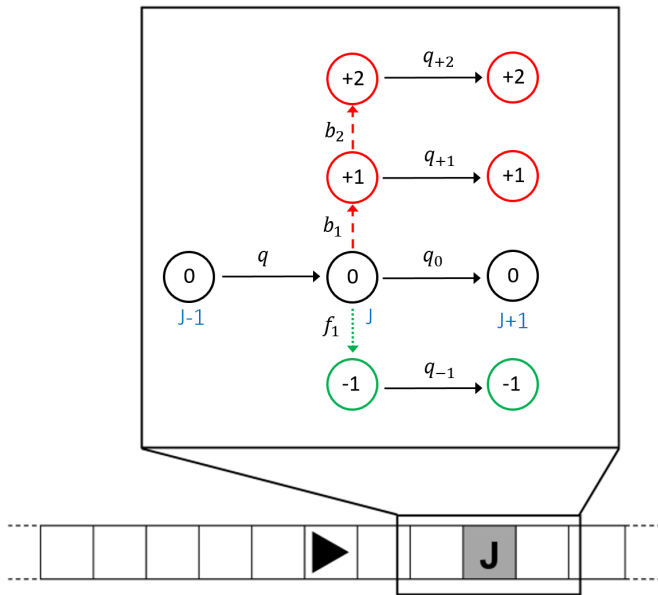
FIG. 2. A kinetic model for the transcription slippage at site $J$ on the DNA template. The conformational states of RNAP during transcription with first backward slippage of nascent mRNA transcript, transcription with second backward slippage of nascent mRNA transcript, transcription with forward slippage of nascent mRNA transcript, and normal transcription are represented by $+1$, $+2$, $-1$, and $0$, respectively. $b_1$, $b_2$, and $f_1$ are the first backward slippage rate (dashed arrow), second backward slippage rate (dashed arrow), and first forward slippage rate (dotted arrow), respectively. $q$, $q_0$, $q_{+1}$, $q_{-1}$, and $q_{+2}$ are the normal transcription rates (simple arrows).

## A. TS phenomenon

During normal transcription [see Fig. 1(a)], incorporation of every nucleotide, as directed by the template DNA, is followed by the motorlike forward movement of the RNAP by one nucleotide along the DNA that also serves as its track. Thus, in normal transcription, each event of elongation of the nascent RNA transcript by one nucleotide is tightly coupled to the translocation of RNAP by one nucleotide on the template DNA. A common slippery sequence on the DNA is a sequence of A's [5] (recall "A" is one of the four different types of subunits of DNA). Figure 1(b) describes a transcription process where a single backward slippage of the nascent RNA transcript occurs. Due to the backward slippage, the active site of the RNAP turns empty for the second time during its sojourn at the slippery site, and it transcribes the same nucleotide for a second time, resulting in the insertion of an extra nucleotide on the transcript. Figure 1(c) describes a single forward slippage of the nascent RNA transcript without forward movement of RNAP; this forward slippage of the nascent transcript results in the active site of the RNAP getting occupied with the previously added nucleotide in the transcript thereby preventing incorporation of a fresh nucleotide even though the RNAP moves forward by one nucleotide on the DNA template.

It is worth emphasizing that the slippage phenomena described above do not imply any movement of the RNAP with respect to its DNA track. Instead, in this paper we are concerned with the slippage of nascent RNA with respect to the RNAP as well as the DNA template. In both Figs. 1(b) and 1(c), the transcript slips without the concomitant movement of RNAP and the polymerization of the transcript is not coupled with the translocation of RNAP. Typically, either of these mechanisms can be repeated thereby causing multiple rounds of backward or forward slippage of the transcript at these slippery sites.

## B. Kinetic model motivated by TS

In our model, we represent the DNA template as a one-dimensional lattice of length $L$. We label the sites of the lattice by the integer index $j$ $(1 \leqslant j \leqslant L)$. Each lattice site corresponds to a nucleotide on the DNA template. The instantaneous position of a RNAP is denoted by the integer index $j$; in each round of successful error-free elongation of the nascent RNA by one unit, the RNAP takes a forward step from $j$ to $j+1$. The special site where TS can take place has been labeled by the integer $J$ (i.e., $j = J$). Since our study is primarily on TS, and since TS is known [5] to occur at a special slippery site, we focus in this section exclusively on the triplet of sites $J-1$, $J$, and $J+1$.

For a completely normal error-free transcription of the full-length template DNA, the RNAP takes $L$ steps on the track synthesizing a RNA transcript of length $L$, i.e., exactly equal to the length of the DNA template. However, in the case of transcription with $n$ successive rounds of backward slippage at a specially designated slippery site, insertion of $n$ number of nucleotides leads to the synthesis of a longer transcript of total length $L+n$. Similarly, for $n$ successive rounds of forward slippage at the slippery site, missing the transcription of $n$ nucleotides on the template (i.e., effectively, deletion of $n$ nucleotides) produces a shorter transcript of total length $L-n$. Backward slippages have been found to occur more often than the forward slippage.

The extra length of the nascent RNA caused by the slippage is labeled by an integer index $\mu$ that can, in principle, be positive, negative, or zero. According to our convention $\mu$ is positive (negative) in the case of backward (forward) slippage; in contrast, $\mu = 0$ if the nascent transcript suffers no slippage or it suffers equal numbers of forward and backward slippages at the slippery site $J$. In other words, $\mu$ denotes the slippage-induced length change of the product transcript as compared to that of the template. Throughout this paper we use the term "slippage state" to denote the magnitude of $\mu$.

The theoretical framework that we have formulated is very general and can treat any arbitrary number $N_b$ of backward or $N_f$ number of forward slippage of the transcript while the RNAP is occupying the specific lattice site $J$. However, for the purpose of presentation of concrete results here through an explicit analytical calculation, we have allowed a maximum of two backward slippage events ($N_b = 2$, corresponding to $\mu = +1, +2$), and a maximum of a single forward slippage ($N_f = 1$ that would correspond to $\mu = -1$). Since backward slippages have been found to occur more often than the reverse process, in the example shown in Fig. 2, possibilities of two successive backward slippages are shown against the possibility of a single forward slippage. For the same reason, we exclude the possibility of a backward slippage followed by a

forward slippage which could result in a RNA of normal size. Moreover, allowing a larger number of slippage (i.e., larger value of $|\mu|$) would create a jam in RNAP traffic because at the slippery site each RNAP would have to wait much longer because of the larger number of slippages of its nascent mRNA due to the higher allowed value of $|\mu|$.

The allowed transitions are indicated by the arrows and the corresponding rates are also shown next to the respective arrows in Fig. 2. The special case shown in Fig. 2 is, indeed, simple enough to be treated analytically. However, analytical calculations become more and more difficult with the increasing number of backward (or forward) slippage events (i.e., with the increase in the allowed values of $N_b$ and $N_f$). Nevertheless, in principle, the strategy of modeling followed here can be implemented numerically also for any arbitrary values of $N_b$ and $N_f$ if calculations become too difficult to carry out analytically.

Normal transcription, without any slippage, at the special site $J$ would correspond to the transition $(J, 0) \rightarrow (J + 1, 0)$. In contrast, a backward TS at the site $J$ results in the transition $(J, 0) \rightarrow (J, +1)$. Therefore, a single backward TS followed by a normal transcription together result in the composite transition $(J, 0) \rightarrow (J + 1, +1)$. Since no TS are allowed to occur at any site before or after the special site $J$, the value of $\mu$ attained finally at $J$ is carried throughout the subsequent transit of the RNAP until the termination of transcription at the site $i = L$.

Thus, for the special case of the model shown in Fig. 2, the state of the RNAP motor at a particular instant is indicated by the pair $(j, \mu)$, where $j$ is its position on the DNA template, i.e., $j = 1$ to $L$ and $\mu$ ($\mu = 0, +1, +2, -1$) is the "extra length" of the associated nascent transcript. Figure 2 clearly shows that, in this special case, a RNAP can follow four different pathways when it encounters a slippery site: (1) the transitions $(J - 1, 0) \rightarrow (J, 0) \rightarrow (J + 1, 0)$ correspond to the normal transcription at the slippery site, (2) the transitions $(J - 1, 0) \rightarrow (J, 0) \rightarrow (J, +1) \rightarrow (J + 1, +1)$ correspond to the single backward slippage of transcript at the slippery site, (3) the transitions $(J - 1, 0) \rightarrow (J, 0) \rightarrow (J, +1) \rightarrow (J, +2) \rightarrow (J + 1, +2)$ correspond to the double backward slippage of transcript at the slippery site, and (4) the transitions $(J - 1, 0) \rightarrow (J, 0) \rightarrow (J, -1) \rightarrow (J + 1, -1)$ correspond to

the single forward slippage of transcript at the slippery site.

## III. PASSAGE OF RNAP ACROSS SLIPPERY SITE SUFFERING TRANSCRIPT SLIPPAGE

### A. First-passage times across slippery site and transient behavior

We define the time $\tau$ taken by the RNAP motor to reach, *for the first time*, the position $J + 1$, starting from the position $J - 1$ as the *first-passage time*. Since the kinetics of transcription, including TS, is probabilistic, $\tau$ varies from one RNAP to another. In this section we calculate the probability distribution (more precisely, the probability density function) $f(\tau)$ of $\tau$. The mean first-passage time can be obtained from

$$\langle \tau \rangle = \int_0^\infty \tau f(\tau) d\tau \tag{1}$$

if $f(\tau)$ is known.

We define $P_\mu(J, t)$ as the probability of finding the RNAP in the "slippage state" $\mu$ at site $J$ on the DNA track at time $t$. The master equations governing the time evolution of $P_\mu(J, t)$ corresponding to the general $N$-state kinetic model, written using the matrix notation, is

$$\frac{d\mathbf{P}(t)}{dt} = A\mathbf{P}(t), \tag{2}$$

where the vector $\mathbf{P}(t)$ is an $N$-component column vector, the components of which are $P_0(J - 1, t)$, $P_0(J + 1, t)$, and $P_\mu(J, t)$ ($\mu = 0, +1, +2, -1$), and the elements of the matrix $A$ are the rates of transitions (more appropriately, the transition probabilities per unit time) between these states. For example, in the special case of the six-state kinetic model shown in Fig. 2, defining

$$\mathbf{P}(t) = \begin{pmatrix} P_0(J - 1, t) \\ P_0(J, t) \\ P_0(J + 1, t) \\ P_{+1}(J, t) \\ P_{-1}(J, t) \\ P_{+2}(J, t) \end{pmatrix} \tag{3}$$

we have

$$A = \begin{pmatrix} -q & 0 & 0 & 0 & 0 & 0 \\ q & -(q_0 + b_1 + f_1) & 0 & 0 & 0 & 0 \\ 0 & q_0 & 0 & q_{+1} & q_{-1} & q_{+2} \\ 0 & b_1 & 0 & -(q_{+1} + b_2) & 0 & 0 \\ 0 & f_1 & 0 & 0 & -q_{-1} & 0 \\ 0 & 0 & 0 & b_2 & 0 & -q_{+2} \end{pmatrix}. \tag{4}$$

Sometimes the computation of the distribution of FPT turns out to be easier in terms of Laplace transforms. Carrying out Laplace transform of both sides of Eq. (2),

$$\mathcal{L}\left[\frac{d\mathbf{P}(t)}{dt}\right] = A\mathcal{L}\mathbf{P}(t), \tag{5}$$

we get

$$s\tilde{\mathbf{P}}(s) - \mathbf{P}(0) = A\tilde{\mathbf{P}}(s) \tag{6}$$

and hence

$$\tilde{\mathbf{P}}(s) = (s\mathbf{I} - A)^{-1}\mathbf{P}(0), \tag{7}$$

where $\mathbf{I}$ is the identity matrix, $\mathcal{L}$ indicates the Laplace transform operator, and $\tilde{\mathbf{P}}(s)$ is the Laplace transform of $\mathbf{P}(t)$. In principle, after taking the inverse Laplace transform of $\tilde{\mathbf{P}}(s)$, one would get the distribution of FPT:

$$\mathbf{P}(t) = \mathcal{L}^{-1}[\tilde{\mathbf{P}}(s)]. \tag{8}$$

Often the set of kinetic equations is so complicated that the operation of inverse Laplace transform (8) cannot be completed analytically to get a closed-form analytical expression for $P_\mu(J, t)$. In such situations, the mean first-passage time can still be obtained by taking appropriate derivatives of $\tilde{P}_\mu(J, s)$ if the latter can be calculated in the $s$ space (Laplace space):

$$\int_0^\infty t P_\mu(J, t) dt = -\frac{d}{ds} \tilde{P}_\mu(J, s)\Big|_{s=0} \tag{9}$$

along with the normalization condition

$$P_0(J - 1, t) + \sum_\mu P_\mu(J, t) + P_0(J + 1, t) = 1. \tag{10}$$

For the calculation of the first-passage time, we impose the initial conditions:

$$\mathbf{P}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{11}$$

The probability density of first-passage times to reach the target site $J + 1$ between time $t$ and $t + dt$ is

$$f(t) = q_0 P_0(J, t) + q_{+1} P_{+1}(J, t) + q_{-1} P_{-1}(J, t)$$
$$+ q_{+2} P_{+2}(J, t). \tag{12}$$

The transition matrix $A$ (4) can be simply diagonalized as follows:

$$A|\lambda_n\rangle = \lambda_n |\lambda_n\rangle, \tag{13}$$

where the eigenvalues $\lambda_n$ are given by

$$\lambda_1 = 0, \quad \lambda_2 = -q, \quad \lambda_3 = -b_1 - f_1 - q_0,$$
$$\lambda_4 = -q_{-1}, \quad \lambda_5 = -b_2 - q_{+1}, \quad \lambda_6 = -q_{+2}. \tag{14}$$

The time dependence of the eigenvectors is simply written as

$$|\lambda_n(t)\rangle = e^{\lambda_n t}|\lambda_n\rangle. \tag{15}$$

Now we expand the initial state $|P(0)\rangle$, given by (11), in terms of the eigenvectors of $A$ as

$$|P(0)\rangle = \sum_{n=1}^6 c_n |\lambda_n\rangle, \tag{16}$$

where

$$c_1 = 1, \quad c_2 = \frac{b_1 b_2 q}{x_1 x_2 x_3}, \quad c_3 = \frac{b_1 b_2 q}{x_1 x_{21} x_{13}},$$
$$c_4 = \frac{f_1 q}{x_4 x_{41}}, \quad c_5 = \frac{b_1 b_2 q}{x_2 x_{12} x_{23}}, \tag{17}$$
$$c_6 = \frac{b_1 b_2 q}{x_3 x_{13} x_{32}}$$

and

$$x_1 = b_1 + f_1 + q_0 - q, \quad x_2 = b_2 + q_{+1} - q,$$
$$x_3 = q_{+2} - q, \quad x_4 = q_{-1} - q, \tag{18}$$
$$x_{jk} = x_j - x_k,$$

from which we have

$$|P(t)\rangle = \sum_{n=1}^6 c_n e^{\lambda_n t} |\lambda_n\rangle. \tag{19}$$

Explicitly, each component can be written as

$$P_0(J - 1, t) = e^{-qt},$$
$$P_0(J, t) = q e^{-qt}\left(\frac{1 - e^{-x_1 t}}{x_1}\right),$$
$$P_{+1}(J, t) = \frac{b_1 q e^{-qt}}{x_{12}}\left(\frac{1 - e^{-x_2 t}}{x_2} - \frac{1 - e^{-x_1 t}}{x_1}\right),$$
$$P_{-1}(J, t) = \frac{f_1 q e^{-qt}}{x_{14}}\left(\frac{1 - e^{-x_4 t}}{x_4} - \frac{1 - e^{-x_1 t}}{x_1}\right),$$
$$P_{+2}(J, t) = \frac{b_1 b_2 q e^{-qt}}{x_{12} x_{23} x_{13}}$$
$$\times \left(x_{12}\frac{1 - e^{-x_3 t}}{x_3} + x_{23}\frac{1 - e^{-x_1 t}}{x_1}\right.$$
$$\left. - x_{13}\frac{1 - e^{-x_2 t}}{x_2}\right). \tag{20}$$

Finally, the expression for the remaining probability $P_0(J + 1, t)$ can be obtained simply using the normalization condition (10), i.e.,

$$P_0(J + 1, t) = 1 - \{P_0(J - 1, t) + P_0(J, t) + P_{+1}(J, t)$$
$$+ P_{-1}(J, t) + P_{+2}(J, t)\}. \tag{21}$$

The exact expressions (20) and (21) for the six probabilities $P_0(J - 1, t), P_0(J, t), P_0(J + 1, t), P_{+1}(J, t), P_{-1}(J, t)$ and $P_{+2}(J, t)$ are drawn graphically for two sets of the rate constants in Figs. 3(a) and 3(b). On the same graph we also plot the corresponding numerical data obtained from our Monte Carlo simulations of the model with the same set of values of the rate constants. The excellent agreement between the theory and simulation establishes the high accuracy of the simulation data because the analytical expressions (20) are exact. The variation of the probabilities with time are consistent with the intuitive expectation based on the initial conditions. The only difference between Fig. 3(a) and Fig. 3(b) is that $P_{+2}(J, t)$ saturates to a constant value in the latter, whereas it decays to zero in the former. This qualitative difference arises from the choice of parameter value $q_{+2} = 0$ in Fig. 3(b) because of which probability $P_{+2}(J, t)$ cannot decay to zero as $t \to \infty$. If $q_{+2} = 0$, transcription pauses after two backward slippages. It means that there is no leakage of probability after the second backward slippage, so the corresponding probability cannot decay to zero as $t \to \infty$.

The exact expression of the distribution of first-passage times can now be obtained by substituting (20) and (21) into the relation (12). Figure 4 shows the distribution of the *first-passage time* $(\tau)$ for four different values of $b_1$ keeping all the other parameter values fixed.
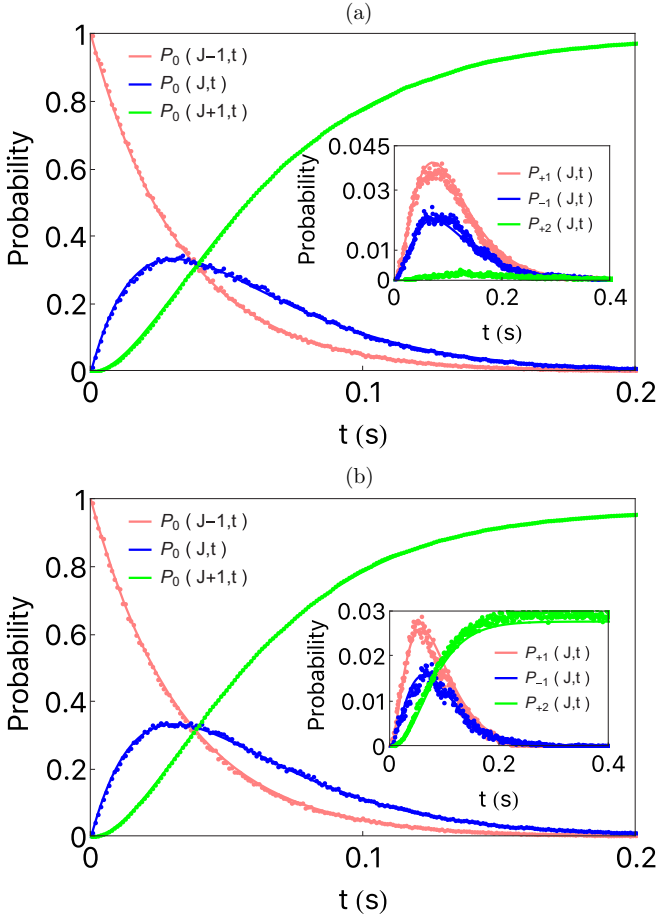
FIG. 3. Distribution of probabilities $\mathbf{P}(t)$ against time $t$; for (a) $q_{+2} = 10$ s$^{-1}$, $b_2 = 1$ s$^{-1}$ and (b) $q_{+2} = 0$ s$^{-1}$, $b_2 = 10$ s$^{-1}$. All the other parameters are kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, $b_1 = 4$ s$^{-1}$, and $f_1 = 2$ s$^{-1}$. Lines correspond to analytical results and discrete data points were obtained from simulation. Inset plot shows how the probabilities of slippage states vary with time.
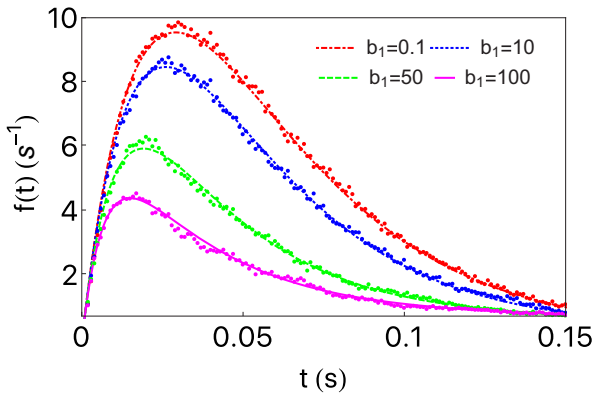


FIG. 4. Distribution of probability density of first-passage time $f(t)$ as a function of time ($t$); for different values of $b_1$. All the other parameters are kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = q_{+2} = 1$ s$^{-1}$, and $b_2 = f_1 = 10$ s$^{-1}$. Lines correspond to analytical results and discrete data points are obtained from simulation.
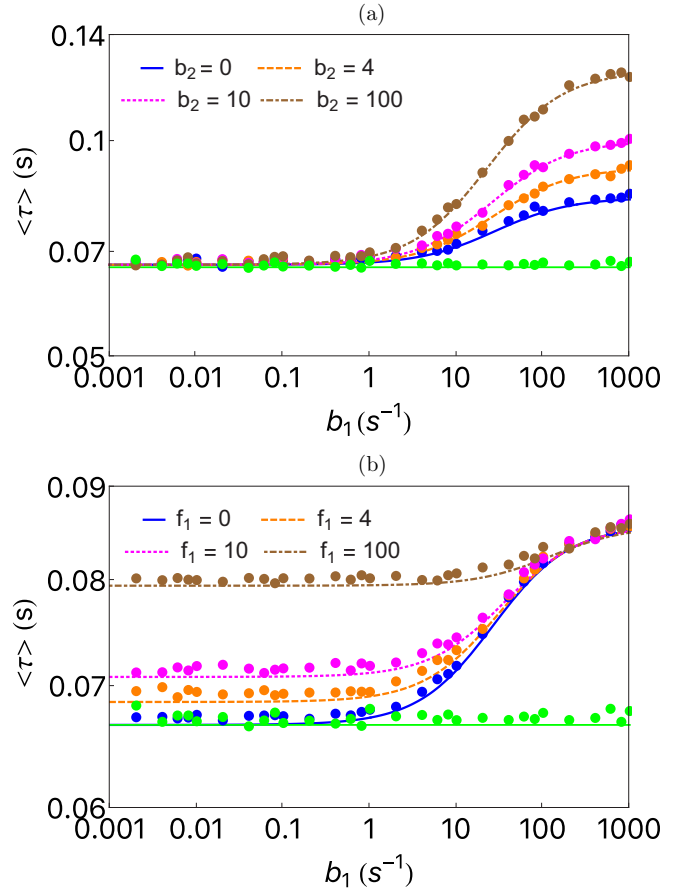


FIG. 5. Variation of $\langle \tau \rangle$ with respect to variation of first backward slippage rate $b_1$; for (a) fixed $f_1 = 1$ s$^{-1}$ and (b) fixed $b_2 = 1$ s$^{-1}$. All other parameters are kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, and $q_{+2} = 10$ s$^{-1}$. Lines correspond to analytical results and points represent the simulation data. The only exception is that the green line (simulation data) corresponds to $q = q_0 = q_{+1} = q_{-1} = q_{+2} = 30$ s$^{-1}$.

The variation of the mean first-passage time $\langle \tau \rangle$ with $b_1$ is shown for four different values of $b_2$ in Fig. 5(a) and for four different values of $f_1$ in Fig. 5(b). The higher the rate of slippage the longer it takes for the RNAP to pass the defect site.

### B. Steady state: Fractions of slipped transcripts

Since no TS is assumed to occur at the $L - 1$ sites labeled by $j \neq J$, the lengths of the transcripts synthesized by the RNAPs in the six-state model can have only the lengths $L$, $L - 1$, $L + 1$, and $L + 2$. For the steady state of the system, we can define the corresponding probabilities by the relations

$$P_L = \frac{P_0(J+1)}{P(J+1)} = \frac{q_0}{b_1 + f_1 + q_0},$$

$$P_{L+1} = \frac{P_{+1}(J+1)}{P(J+1)} = \frac{b_1 q_{+1}}{(b_1 + f_1 + q_0)(b_2 + q_{+1})},$$

$$P_{L-1} = \frac{P_{-1}(J+1)}{P(J+1)} = \frac{f_1}{b_1 + f_1 + q_0},$$

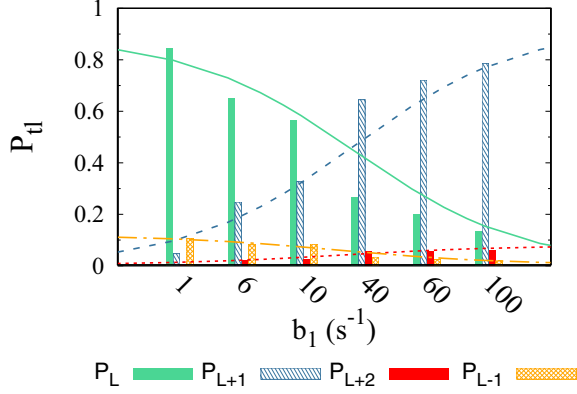$$P_{L+2} = \frac{P_{+2}(J+1)}{P(J+1)} = \frac{b_1 b_2}{(b_1 + f_1 + q_0)(b_2 + q_{+1})}, \quad (22)$$

FIG. 6. Histogram of probability of types of nascent mRNA are plotted with different values of first backward slippage rate $b_1$. Parameter values are $b_2 = 1$ s$^{-1}$ and $f_1 = 2$ s$^{-1}$. All other parameters are kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, and $q_{+2} = 10$ s$^{-1}$. Lines correspond to exact expressions (22), where solid, dashed, dotted, dotdashed lines correspond to $P_L$, $P_{L+1}$, $P_{L+2}$, and $P_{L-1}$, respectively. Bar plots are obtained from simulation data.

where

$$P(J + 1) = P_0(J + 1) + P_{+1}(J + 1) + P_{+2}(J + 1)$$
$$+ P_{-1}(J + 1). \tag{23}$$

Thus, $P_L, P_{L+1}, P_{L-1}, P_{L+2}$ can be interpreted as the fractions of the respective species of the transcripts synthesized. The expressions (22) for the probabilities depend only on the position, not on time $t$, because these correspond to the NESS which is attained in the limit $t \to \infty$.

Figure 6 shows the histogram for probability of transcript lengths ($P_{tl}$) plotted against single backward slippage rate, $b_1$. For low values of $b_1$, $P_L$ is higher than $P_{L+1}$, $P_{L+2}$, and $P_{L-1}$, as the chances of backward slippage are lower than that of forward slippage. For high values of $b_1$, the possibility of backward slippage gets enhanced and hence $P_{L+1}$ is higher than $P_{L+2}$, $P_{L-1}$, and $P_L$. Our analytical results show that
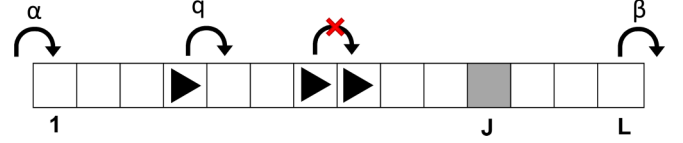


FIG. 7. Schematic diagram of RNAP traffic on the DNA track of length $L$ in the presence of slippage at site $J$. Triangles represent (single walking) RNAPs. RNAPs can enter the DNA track only at site $i = 1$ with rate $\alpha$ if the entry site is empty and RNAPs can leave the DNA track after it reaches the termination site $i = L$ with rate $\beta$. In between, RNAPs can hop forward, if the target site is empty, with the normal transcription rate $q$ except at site $J$, where it can hop with rates $q_\mu$ ($\mu = 0, \pm 1, +2$), depending on the slippage state (refer to Fig. 2).

the probability of $P_{L+1}$ and $P_{L+2}$ increases as $b_1$ increases and $P_L$ and $P_{L-1}$ decreases as $b_1$ increases. The trend in the histogram plot qualitatively matches with recent experimental findings [42,43].

## IV. EFFECT OF RNAP TRAFFIC CONGESTION ON THE TRANSCRIPT SLIPPAGE PHENOMENON

Figure 7 shows a schematic diagram of the model for RNAP traffic on the DNA track in the presence of a slippery site at $J$. A RNAP can enter the DNA track, with rate $\alpha$, if the entry site $i = 1$ is empty. If the RNAP is located at any other position $i \neq J$, it can move forward, with rate $q$ if, and only if, the target site is empty. On the other hand, while located at the special site $J$ a RNAP can hop forward with rates $q_\mu$ ($\mu = 0, \pm 1, +2$), depending on the slippage state (refer to Fig. 2). A RNAP can detach from the track at the exit site $i = L$ with rate $\beta$. At the slippery site $J$, the nascent RNA can slip backward with rates $b_1$, $b_2$, etc., and forward with the rate $f_1$. Since TS does not involve any movement of the RNAP with respect to its DNA track, forward slippage (of the RNA transcript) can happen even when the next site in front of the RNAP is covered by another RNAP.

Let $P_\mu(i, t)$ denote the probability of finding RNAP in slippage state $\mu$ at site $i$ on the DNA track at time $t$. So, the probability that the site $i$ is occupied by a RNAP at time $t$, irrespective of its slippage state, is $P(i, t) = \sum_\mu P_\mu(i, t)$, where $\mu = 0, +1, +2, -1$. We can refer to this model as a biologically motivated extension of the TASEP with a special kind of defect located at the specific site $i = J = L/2$. Under mean-field approximation, the master equation for the probabilities $P_\mu(i, t)$ are given by

$$\frac{dP(1, t)}{dt} = \alpha[1 - P(1, t)] - qP(1, t)[1 - P(2, t)],$$

$$\frac{dP(i, t)}{dt} = q\{P(i - 1, t)[1 - P(i, t)] - P(i, t)[1 - P(i + 1, t)]\} \qquad \text{for } 1 < i < L \ (i \neq L/2, L/2 + 1),$$

$$\frac{dP_0(L/2, t)}{dt} = qP(L/2 - 1, t)[1 - P(L/2, t)] - q_0 P_0(L/2, t)[1 - P(L/2 + 1, t)] - (b_1 + f_1)P_0(L/2, t),$$

$$\frac{dP_{+1}(L/2, t)}{dt} = b_1 P_0(L/2, t) - q_{+1} P_{+1}(L/2, t)[1 - P(L/2 + 1, t)] - b_2 P_{+1}(L/2, t),$$

$$\frac{dP_{+2}(L/2, t)}{dt} = b_2 P_{+1}(L/2, t) - q_{+2} P_{+2}(L/2, t)[1 - P(L/2 + 1, t)],$$

$$\frac{dP_{-1}(L/2, t)}{dt} = f_1 P_0(L/2, t) - q_{-1} P_{-1}(L/2, t)[1 - P(L/2 + 1, t)],$$

$$\frac{dP(L/2+1,t)}{dt} = [q_0 P_0(L/2,t) + q_{+1}P_{+1}(L/2,t) + q_{+2}P_{+2}(L/2,t) + q_{-1}P_{-1}(L/2,t)][1 - P(L/2+1,t)]$$

$$- qP(L/2+1,t)[1 - P(L/2+2,t)], \quad \frac{dP(L,t)}{dt} = qP(L-1,t)[1 - P(L,t)] - \beta P(L,t). \quad (24)$$

In the steady state the left-hand sides of all these equations vanish and the corresponding solutions of the equations are obtained iteratively by checking whether the difference of the numerical values of two successive iterations is less than $\epsilon$ where $\epsilon \approx 10^{-8}$ is a preassigned small number.

In our Monte Carlo simulation of the model, starting from an initial condition, the flux was monitored in each run to ensure that the system reaches a steady state where the flux becomes independent of time. Then, starting from an even longer instant of time $t_{\text{steady}}$ ($\equiv n_{\text{steady}} dt$) the numerical data from the simulation were recorded for the computation of the steady-state properties; the collection of the data was continued until a time $t_{\text{max}}$ ($\equiv n_{\text{max}} dt$) where the simulation run was terminated. The symbols $n_{\text{steady}}$ and $n_{\text{max}}$ refer to the corresponding number of Monte Carlo steps and each Monte Carlo step corresponds to the infinitesimal real time interval $dt$. The choice of the precise value of $dt$ is based on the fastest rate in the model as the corresponding probability of occurrence of the event must always be less than 1. In our Monte Carlo simulations we chose the numerical value $dt = 5 \times 10^{-4}$ s.

For the correspondence between the results obtained for arbitrary rates $\alpha$ and $\beta$ (with dimensions of inverse time) and the well-known standard results for the TASEP in terms of dimensionless probabilities, where the hopping probability is taken to be unity, requires dividing the rates $\alpha$ and $\beta$ by the rate $q$. In other words, the correspondence requires $\alpha \to \alpha' = \alpha/q$ and $\beta \to \beta' = \beta/q$. In low density (LD) phase, density is determined by $\alpha'$ ($\rho = \alpha'$) and hence $\alpha' < 1/2 < \beta'$. In high density (HD) phase, density is determined by $\beta'$ ($\rho = 1 - \beta'$) and hence $\alpha' > 1/2 > \beta'$. In maximal current (MC) phase, $\rho = 1/2$ and hence $\alpha' = \beta' > 1/2$.

### Effects of traffic congestion on extent of TS

In this section, we compute the mean time taken by each RNAP to transcribe a DNA template of length $L$, on which the defect (i.e., the slippery site) is located at $J$, in the steady state of the RNAP traffic. The simplest way to account for the traffic congestion is to replace the hopping rates of RNAPs, from one site to the next, by effective rates obtained by multiplying the actual rate with the factor $(1 - \rho)$, where $\rho$ is the number density of the RNAPs.

Figure 8 shows the variation in probability distributions of first backward slippage state plotted against time for different values of $\rho$. The trend observed in the graph is due to crowding, i.e, the RNAPs have to face hindrance to move forward.

The denser the traffic congestion, the longer the dwell time of an arbitrary RNAP at the slippery site and the larger the expected number of TS events that can occur during the

duration of that dwell. This intuitive expectation is, indeed, supported by the data shown in Fig. 9 where $P_{+2}(J)$ and $P_{+2}(J+1)$ have been plotted as functions of the number density $\rho$.

For plotting this figure we have chosen $b_1 = f_1 = q_{+2} = 0.3$ s$^{-1}$ and $q = q_0 = q_1 = q_{-1} = b_2 = 30$ s$^{-1}$. Because of the small values of $b_1$ and $f_1$, the likelihood of the first TS event, irrespective of forward or backward, is normally quite low. However, as the value of $\rho$ increases, the dwell times of the RNAPs increase at all sites, including that located at the slippery site. Consequently, during that longer period of stay at the slippery site, the RNAP suffers multiple rounds of TS; this is reflected in the increase in the magnitude of $P_{+2}(J)$ in Fig. 9.

The probabilities $P_{+2}(J)$ and $P_{+2}(J+1)$ are plotted as functions of $\alpha$ (for fixed $\beta$) in Fig. 10(a) and as a function of $\beta$ (for fixed $\alpha$) in Fig. 10(b). As $\alpha$ increases both $P_{+2}(J)$ and $P_{+2}(J+1)$ increase but the rate of increase decreases gradually and, probabilities eventually saturate because the RNAP traffic makes a transition from the LD phase to the MC phase where the flux of RNAPs saturates. Similarly, for a fixed $\alpha$, as $\beta$ increases the transition from the HD phase to MC reduces the effective dwell time of each RNAP at the defect site which, in turn, reduces the probabilities of multiple TS events at that site. Moreover, the transition to the MC phase also leads to the saturations of the values of $P_{+2}(J)$ and $P_{+2}(J+1)$ with increasing $\beta$.

Figure 11 shows the histogram for slippage statistics plotted against $\rho$. In the MC phase ($\rho = 0.5$) and even in the LD phase ($\rho < 0.5$), $P_{L+1}$ and $P_{L-1}$ have significantly low values
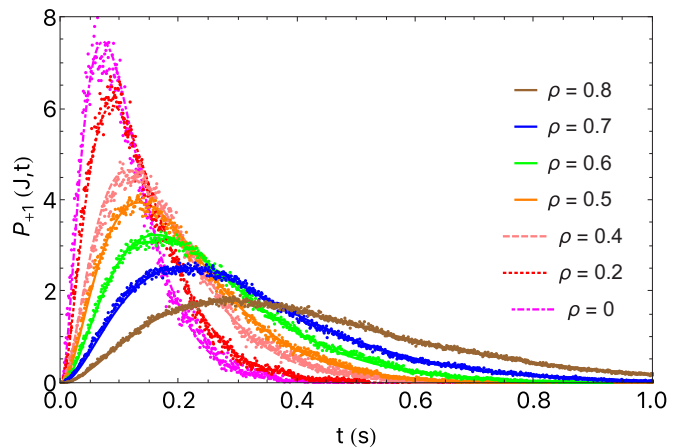
FIG. 8. Distributions of probability of first backward slippage state, $P_{+1}(J,t)$ with time $t$, for different values of $\rho$. Parameter values for slippage rates: $b_1 = 4$ s$^{-1}$, $b_2 = 1$ s$^{-1}$, and $f_1 = 2$ s$^{-1}$. All other parameters were kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, and $q_{+2} = 10$ s$^{-1}$. Lines correspond to analytical results and discrete data points were obtained from simulation.

and remain unaffected by the change in $\rho$. In the HD phase ($\rho > 0.5$), $P_L$ decreases and $P_{L+1}$, $P_{L+2}$, and $P_{L-1}$ increase due to the crowding effect. Because of hindrance, RNAPs have to wait a longer time at the slippage site and it enhances the chance of slippages.

## V. EFFECT OF TRANSCRIPT SLIPPAGE ON RNAP TRAFFIC FLOW

The time taken by a RNAP, on average, in the steady state of the RNAP traffic to transcribe can be written as the inverse of the exit rate $\beta$ multiplied by $P_0(L) + P_{+1}(L) + P_{+2}(L) + P_{-1}(L)$. So,

$$T_{ss} = \frac{1}{\beta P(L)}$$

$$= \frac{1}{\beta[P_0(L) + P_{+1}(L) + P_{+2}(L) + P_{-1}(L)]}. \quad (25)$$

We obtained the mean-field theoretic estimate of $T_{ss}$ by substituting the mean-field values of the probabilities in the denominator of (25).

For the computation of $T_{ss}$ in our Monte Carlo simulation, we use the formula

$$T_{ss} = \frac{(n_{max} - n_{steady})dt}{N}, \quad (26)$$

where $N$ is the total number of departing RNAPs counted at $i = L$ over the (update) step numbers $n_{max} - n_{steady}$ of the simulation and $dt$ is the duration of each time step of the simulation. We have taken $\alpha = \beta = 30$ s$^{-1}$ (maximal current phase) and $dt = 5 \times 10^{-4}$ s = time step for Monte Carlo simulation and MF approximation. We have taken the length of the DNA track ($L$) to be of 1000 lattice sites.

In Fig. 12, we have plotted *steady-state mean time* ($T_{ss}$) as a function of $b_1$, for several different values of $b_2$. The trend of variations of the curves are very similar to those in Fig. 5(a) except for the magnitude of the average times which in Fig. 12 is much longer because of the traffic congestion. The deviation of the theoretical predictions from the numerical simulations
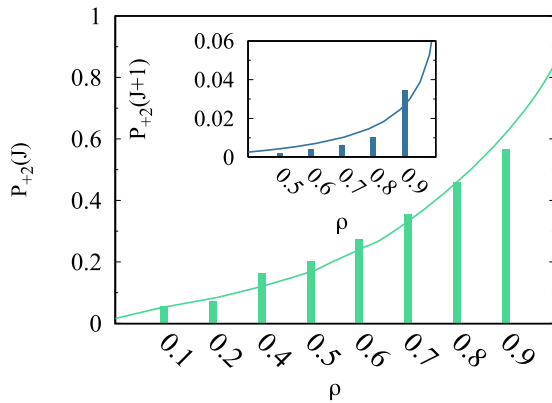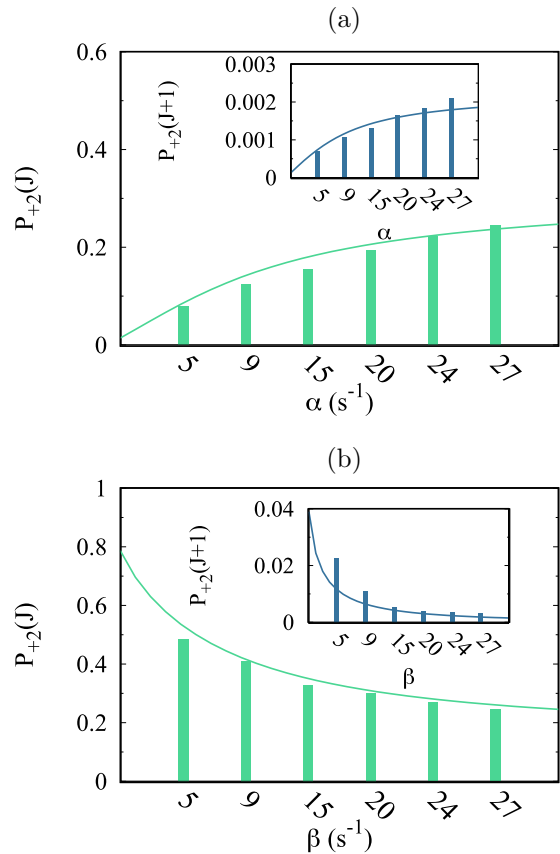
FIG. 10. The steady-state probabilities $P_{+2}(J)$ and $P_{+2}(J + 1)$ are plotted in the main figure and inset, respectively, as functions of entry rate $\alpha$ of RNAP [in (a)] and exit rate $\beta$ of RNAP [in (b)]. For both the figures the parameter values chosen are $b_1 = f_1 = q_{+2} = 0.3$ s$^{-1}$ and $q = q_0 = q_{+1} = q_{-1} = b_2 = 30$ s$^{-1}$; $\beta = 30$ s$^{-1}$ in (a) and $\alpha = 30$ s$^{-1}$ in (b). Lines correspond to MF results and bar plots were obtained from simulation data.

FIG. 9. The steady-state probabilities $P_{+2}(J)$ and $P_{+2}(J + 1)$ are plotted in the main figure and inset, respectively, as functions of the number density $\rho$. The parameter values chosen for this figure are $b_1 = f_1 = q_{+2} = 0.3$ s$^{-1}$ and $q = q_0 = q_{+1} = q_{-1} = b_2 = 30$ s$^{-1}$. Lines correspond to MF results and bar plots were obtained from simulation data.
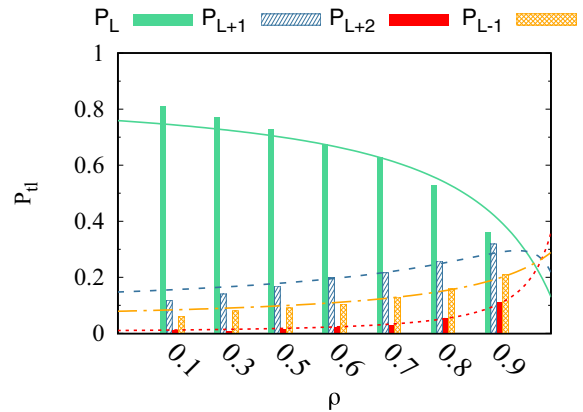
FIG. 11. Histogram of slippage statistics is plotted with different values of RNAP density ($\rho$). Parameter values for slippage rates: $b_1 = 4$ s$^{-1}$, $b_2 = 1$ s$^{-1}$, and $f_1 = 2$ s$^{-1}$. All other parameters are kept fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$ and $q_{+2} = 10$ s$^{-1}$. Lines correspond to the MF result, where solid, dashed, dotted, dotdashed lines correspond to $P_L$, $P_{L+1}$, $P_{L+2}$, and $P_{L-1}$, respectively. Bar plots were obtained from simulation data.
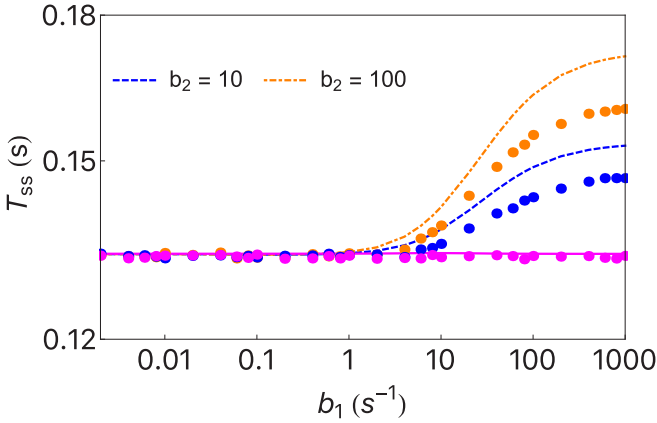
FIG. 12. Variation of the mean transcription time in steady state ($T_{ss}$) with $b_1$ keeping all the other parameters fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, $q_{+2} = 10$ s$^{-1}$, and $f_1 = 1$ s$^{-1}$; $\alpha = \beta = 30$ s$^{-1}$ (maximal current phase) and $dt = 5 \times 10^{-4}$ s = time step for Monte Carlo simulation and MF approximation. Lines correspond to MF theory and points have been obtained from Monte Carlo simulations. The only exception is that the results shown in magenta correspond to parameter values $q = q_0 = q_{+1} = q_{-1} = q_{+2} = 30$ s$^{-1}$.

data in Fig. 12 is mainly because of traffic congestion. In MC phase, mean-field (MF) results are known to deviate significantly from simulation results.

*Steady-state* density profiles of the RNAP traffic in different phases are shown in Fig. 13. We have taken the slippage site to be located at the midpoint of the track ($i = L/2$). This site can be regarded as a defect site on the homogeneous lattice. From the density profile plots, it is clear that the system behaves like a combination of two lattices (or two TASEPs). The TASEP 1 starts at site $i = 1$, where particles can attach with attachment rate $\alpha$, and ends at site $i = L/2$ from where particles can detach with detachment rate $\beta_{\mathrm{eff}}$; in between $i = 1$ and $i = L/2$, particles can hop forward with rate $q$. Similarly, the TASEP 2 starts at site $i = L/2 + 1$, where particles can attach with attachment rate $\alpha_{\mathrm{eff}}$, and ends at site $i = L$ from where particles can detach with detachment rate $\beta$; in between the particles can hop forward with rate $q$.

The LD and HD phases of the TASEP are further divided into the respective subphases LD I, LD II and HD I, HD II which are characterized by the nature of the decay of the density profiles in the boundary layers. In the LD I phase of the TASEP the density profile at the right boundary has positive curvature, whereas in the LD II phase of the TASEP the density profile at the right boundary has negative curvature. Similarly, in the HD I phase of the TASEP the density profile at the left boundary has negative curvature, whereas in the HD II phase of the TASEP the density profile at the left boundary has positive curvature.

For the TASEP without slippage site, the LD phase exists between $\alpha < q/2 < \beta$. With slippage site, the TASEP 1 reaches the LD I phase for $\alpha < q/2$, $\beta_{\mathrm{eff}} < q/2$ (as the relation between the $\alpha$ and $\beta_{\mathrm{eff}}$ does not matter) and the TASEP 2 reaches the LD II phase for $\alpha_{\mathrm{eff}} < q/2 < \beta$ [see in Fig. 13(a)]. Similarly for the TASEP without slippage site, the HD phase exists between $\alpha > q/2 > \beta$. With slippage site,
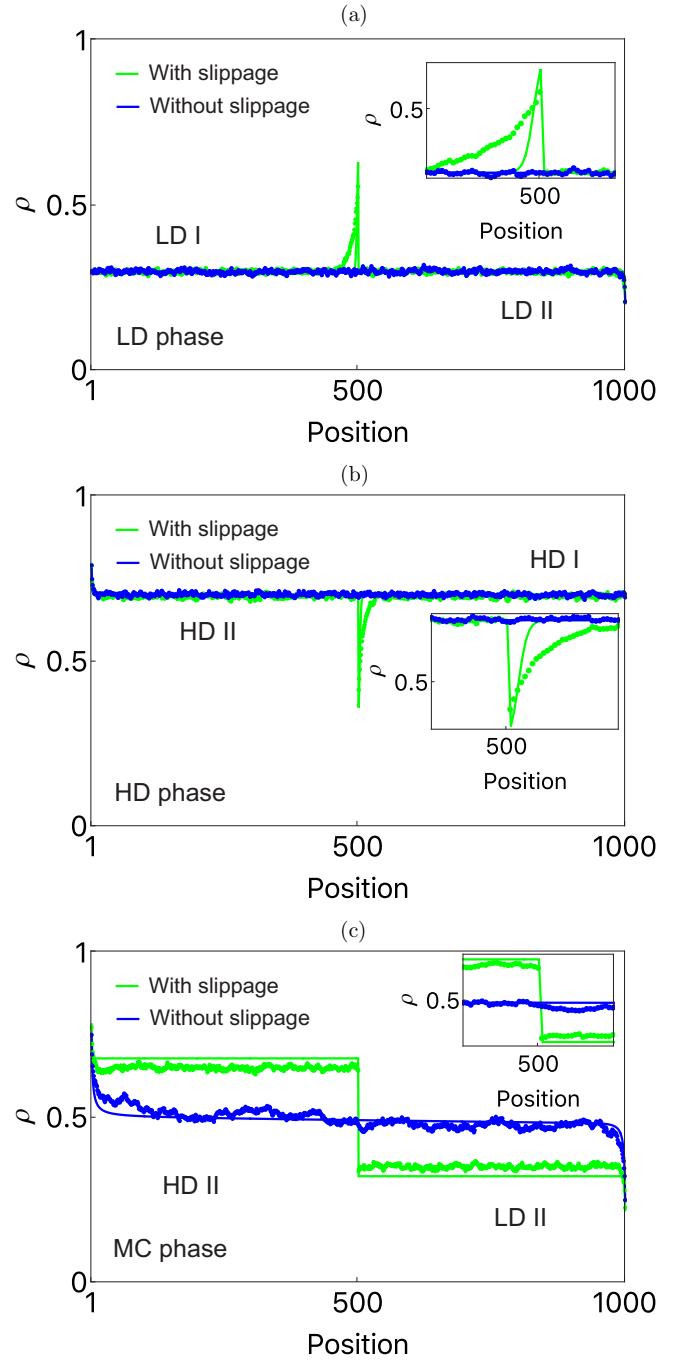






FIG. 13. *Steady-state* density profile of RNAP traffic for (a) LD phase, (b) HD phase, and (c) MC phase. For LD: $\alpha = 9$ s$^{-1}$, $\beta = 30$ s$^{-1}$; HD: $\alpha = 30$ s$^{-1}$, $\beta = 9$ s$^{-1}$; and MC: $\alpha = 30$ s$^{-1}$, $\beta = 30$ s$^{-1}$. All other parameters remain fixed at values $q = q_0 = 30$ s$^{-1}$, $q_{+1} = q_{-1} = 20$ s$^{-1}$, $q_{+2} = 10$ s$^{-1}$, $b_1 = 1000$ s$^{-1}$, $b_2 = 10$ s$^{-1}$, and $f_1 = 1$ s$^{-1}$. Lines correspond to MF theory and points have been obtained from Monte Carlo simulations.

the TASEP 1 reaches the HD II phase for $\alpha > q/2 > \beta_{\mathrm{eff}}$ and the TASEP 2 reaches the HD I phase for $\alpha_{\mathrm{eff}} < q/2$, $\beta < q/2$ (as the relation between the $\alpha_{\mathrm{eff}}$ and $\beta$ does not matter) [see in Fig. 13(b)]. Similarly, for the TASEP without slippage site, the MC phase exists between $\alpha > q/2$, $\beta > q/2$. With slippage site, the MC phase disappears and the TASEP 1 reaches

the HD II phase for $\alpha > q/2 > \beta_{\text{eff}}$ and the TASEP 2 reaches the LD II phase for $\alpha_{\text{eff}} < q/2 < \beta$ [see in Fig. 13(c)].

## VI. SUMMARY AND CONCLUSION

Motivated by the biological phenomenon of TS [5] in RNAP traffic, we have developed a stochastic kinetic model based on the TASEP where a special lattice site is treated as an unusual "defect." The state of each particle, which represents a RNAP, is denoted by two integer indices. The first index denotes its position on the lattice while the second expresses the extra length of the associated RNA transcript because of TS.

In the first part of this paper, we have derived an exact analytical expression for the mean time taken by a single RNAP to traverse the defect site, in the absence of steric hindrance from any other RNAP. This mean time is extracted from the corresponding probability density distribution that we have derived here using the formalisms of first-passage time. The exact analytical expressions that we report reflect important statistical properties that characterize the passage of a single RNAP across the defect site while motoring along its DNA track.

In the second part of this paper, we have investigated the interplay of TS at the defect site and RNAP traffic on the lattice where the RNAP traffic has been modeled as a TASEP. We have presented multiple evidences to establish an increase in the number of TS events suffered by a RNAP while dwelling at the defect site for longer duration because of the traffic congestion. We have also indicated how the TS process affects the flux of the RNAP traffic. We have found good agreement between our theoretical predictions, based on an approximate analysis of the TASEP model and the corresponding data obtained by carrying out Monte Carlo simulations of the same model.

Our model is very general. However, our analytical treatment of the model is based on the assumption of homogeneity of the DNA sequence. In the realistic case, the transcription rate for every DNA nucleotide will vary depending on the identity of the nucleotide on the template and the concentration of the corresponding free monomers available in the surrounding medium. For quantitative predictions, that can be compared with experimental data, inhomogeneous sequences have to be considered. But, in those cases the analysis can be carried out only numerically because analytical treatment would be very difficult (if not impossible).

In spite of the simplicity of the special case of the model and the approximations made in its analytical treatment, the theoretically predicted probability distribution of the longer and shorter transcripts qualitatively matches the experimental data [42,43] obtained through advanced sequencing technologies [44]. However, for detailed quantitative predictions for specific systems, numerical values of the slippage rates would be required; but, at present, estimates of these rates are not available in the literature. We hope that our theoretical predictions will encourage more experimental studies.

The work reported here is of interdisciplinary nature. It has been motivated by a biological phenomenon, namely, TS. The model proposed for studying this phenomenon is an extension of one of the simplest mathematical models of a system of interacting self-driven particles, namely, the TASEP. The model has been analyzed from the perspective of nonequilibrium statistical physics. A single RNAP operates as a "tape-copying Turing machine" [45–47]; a Turing machine [48] is an idealized device conceptualized for abstract "computation" [49]. Therefore, the general theoretical framework developed here may be of interest also in the theory of computation.

[1] *RNA Polymerases as Molecular Motors*, edited by H. Buc and T. Strick (The Royal Society of Chemistry, London, 2009).

[2] D. Chowdhury, Phys. Rep. **529**, 1 (2013).

[3] D. Chowdhury, Biophys. J. **104**, 2331 (2013).

[4] A. B. Kolomeisky, *Motor proteins and molecular motors* (CRC Press, Boca Raton, FL, 2015).

[5] *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, edited by J. F. Atkins and R. F. Gesteland (Springer, New York, 2010), Vol. 24.

[6] C. L. Turnbough, Curr. Opin. Microbiol. **14**, 142 (2011).

[7] A. Schadschneider, D. Chowdhury, and K. Nishinari, *Stochastic Transport in Complex Systems: From Molecules to Vehicles* (Elsevier, Amsterdam, 2010).

[8] D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000).

[9] D. Chowdhury, A. Schadschneider, and K. Nishinari, Phys. Life Rev. **2**, 318 (2005).

[10] T. Chou, K. Mallick, and R. K. P. Zia, Rep. Prog. Phys. **74**, 116601 (2011).

[11] C. Appert-Rolland, M. Ebbinghaus, and L. Santen, Phys. Rep. **593**, 1 (2015).

[12] T. Tripathi and D. Chowdhury, Phys. Rev. E **77**, 011921 (2008).

[13] S. Klumpp and T. Hwa, Proc. Natl. Acad. Sci. USA **105**, 18159 (2008).

[14] S. Klumpp, J. Stat. Phys. **142**, 1252 (2011).

[15] M. Sahoo and S. Klumpp, Europhys. Lett. **96**, 60004 (2011).

[16] Y. Ohta, T. Kodama, and S. Ihara, Phys. Rev. E **84**, 041922 (2011).

[17] J. Wang, B. Pfeuty, Q. Thommen, M. C. Romano, and M. Lefranc, Phys. Rev. E **90**, 050701(R) (2014).

[18] V. Belitsky and G. M. Schütz, J. Theor. Bio. **462**, 370 (2019).

[19] B. Derrida, Phys. Rep. **301**, 65 (1998).

[20] G. M. Schütz, *Integrable Stochastic Many-Body Systems*, Phase Transitions and Critical Phenomena Vol. 19, edited by C. Domb and J. L. Lebowitz (Academic, London, 2000).

[21] K. Mallick, Physica A **418**, 17 (2015).

[22] M. R. Evans and T. Hanney, J. Phys. A: Math. Gen. **38**, R195 (2005).

[23] R. Blythe and M. R. Evans, J. Phys. A: Math. Gen. **40**, R333 (2007).

[24] S. A. Janowsky and J. L. Lebowitz, Phys. Rev. A **45**, 618 (1992); J. Stat. Phys. **77**, 35 (1994).

[25] G. Tripathy and M. Barma, Phys. Rev. Lett. **78**, 3039 (1997).

[26] G. Tripathy and M. Barma, Phys. Rev. E **58**, 1911 (1998).

[27] M. Barma, Physica A **372**, 22 (2006).

[28] A. B. Kolomeisky, J. Phys. A: Math. Gen. **31**, 1153 (1998).

[29] R. J. Harris and R. B. Stinchcombe, Phys. Rev. E **70**, 016108 (2004).

[30] R. Juhász and L. Santen, F. Iglói, Phys. Rev. E **74**, 061101 (2006).

[31] M. E. Foulaadvand, A. B. Kolomeisky, and H. Teymouri, Phys. Rev. E **78**, 061116 (2008).

[32] P. Greulich and A. Schadschneider, Physica A **387**, 1972 (2008).

[33] P. Greulich and A. Schadschneider, J. Stat. Mech.: Theory Exp. (2008), P04009.

[34] J. Dong, S. Klumpp, and R. K. P. Zia, Phys. Rev. Lett. **109**, 130602 (2012).

[35] J. Schmidt, V. Popkov, and A. Schadschneider, Europhys. Lett. **110**, 20008 (2015).

[36] I. Dhiman and A. K. Gupta, J. Comput. Phys. **309**, 227 (2016).

[37] I. Dhiman and A. K. Gupta, Phys. Lett. A **380**, 2038 (2016).

[38] B. Mishra and D. Chowdhury, Phys. Rev. E **95**, 062117 (2017).

[39] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).

[40] *First-Passage Phenomena and their Applications*, edited by S. Redner, R. Metzler, and G. Oshanin (World Scientific, Singapore, 2014).

[41] S. Iyer-Biswas and A. Zilman, Adv. Chem. Phys. **160**, 261 (2016).

[42] A. Olspert, B. Y.-W. Chung, J. F. Atkins, J. P. Carr, and A. E. Firth, EMBO Rep. **16**, 995 (2015).

[43] A. Olspert, J. P. Carr, and A. E. Firth, Nucleic Acids Res. **44**, 7618 (2016).

[44] M. Beverly, C. Hagen, and O. Slack, Anal. Bioanal Chem. **410**, 1667 (2018).

[45] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).

[46] R. A. Mooney, I. Artsimovitch, and R. Landick, J. Bacteriol. **180**, 3265 (1998).

[47] A. K. Sharma and D. Chowdhury, Biophys. Rev. Lett. **07**, 135 (2012).

[48] R. P. Feynman, *Feynman Lectures on Computation* (Addison-Wesley, Reading, MA, 1996).

[49] M. L. Minsky, *Computation: Finite and Infinite Machines* (Prentice-Hall, Englewood Cliffs, NJ, 1967).