

A traffic flow model for bio-polymerization processes

Lisa Davis · Tomáš Gedeon[†] · Jakub Gedeon · Jennifer Thorenson

Received: date / Accepted: date

Abstract Bio-polymerization processes like transcription and translation are central to proper function of a cell. The speed at which the bio-polymer grows is affected both by the number of pauses of elongation machinery, as well the number of bio-polymers due to crowding effects. In order to quantify these effects in fast transcribing ribosome genes, we rigorously show that a classical traffic flow model is the limit of a mean occupancy ODE model. We compare the simulation of this model to a stochastic model and evaluate the combined effect of the polymerase density and the existence of pauses on the instantaneous transcription rate of ribosomal genes.

Keywords Conservation law · transcription process · transcriptional pausing · rrn gene

Mathematics Subject Classification (2000) 35L65 · 65M08 · 65M60 · 92D20

1 Introduction

Transcription of DNA to mRNA and translation of mRNA to protein are two key cellular processes. Even with the discovery of the regulatory role of microRNA, the control of transcription and translation is crucial in cellular response to changing environment. On a very basic level these processes consist of the motion of a complex machine along a one-dimensional strand. Both transcription (where a polymerase moves along the DNA, producing mRNA)

[†] Corresponding author
Department of Mathematical Sciences
Montana State University
Bozeman MT 59717-2400
Tel.: (406)994-5359
Fax: (406)994-1789
E-mail: gedeon@math.montana.edu

and translation (where ribosomes translocate along the mRNA producing proteins) involve polymerization of new macromolecules. Cellular transport along the actin filaments and microtubules, which does not involve polymerization, also fits into this category. A bio-polymerization process has roughly four parts: assembly of the machinery at the start of the strand, movement initiation, elongation and termination. In our description below we will concentrate on polymerase (RNAP) motion along the DNA, but the main modeling ideas apply to all the cases above.

The motion of RNAP is not uniform. Single molecule observation using optical traps indicate that the elongating polymerase frequently pauses, and the duration of the pauses is roughly bi-modal (Neuman et al. 2003) with means 1.2 ± 0.1 sec with amplitude 60% and 6 ± 0.4 sec with amplitude 40%. There are two types of transcriptional pauses of RNAP (Landick 2009): 1) backtracking pauses, where RNAP slides backwards and nascent mRNA slips past the active polymerization site, and 2) non-backtracking pauses, described in Kireeva and Kashlev (2009), which involve mis-folding of the active site. The first type of pauses is affected by force applied on the RNAP, while the second type is not. As described by Klumpp and Hwa (2008) (see also Klumpp (2011)), the existence of pauses brings up the possibility that at high transcription initiation rates a paused polymerase can prevent the forward motion of those following it, thus creating traffic jams. While the density of polymerases on most genes is probably not sufficient to produce traffic jams, there are special genes, where the density is very high. In the exponentially growing populations of *Escherichia coli*, the number of ribosomes per cell increase from 6700 at 0.6 doublings per hour to 71,000 at 2.5 doublings per hour, and 80% of those ribosomes are engaged in protein production, while the rest are involved in the process of assembling (Bremer and Dennis 1996). Transcription of the ribosomal rRNA accounts for over half of the total transcriptional activity in *E. coli* (Bremer and Dennis 1996), even though rRNA (*rrn*) operons only account for 0.5 % of the total genome. Thus, under favorable environmental conditions, most of the cell's metabolic capacity is devoted to making ribosomes (Condon et al. 1993). To sustain the high cellular demand on ribosomal RNA synthesis, the density of polymerases on the *rrn* operon is very high. The goal of this paper is to quantify the combined effect of high polymerase density and ubiquitous pauses on the transcription rate of ribosomal RNA using a computationally efficient continuum model.

Elongation models fall into three broad categories. The most detailed treat the polymerase motion as a stochastic ratchet with detailed energy balance at each step along the DNA (von Hippel 1998, Bai et al. 2004, Tadigotla et al. 2006). A simplified class of models have a long tradition in the physics community under the TASEP label (Totally Asymmetric Simple Elongation Process) (Spitzer 1970, Janowsky and Lebowitz 1992, Derrida et al. 1993, Schutz and Domany 1993, Kolomeisky 1998, Zia et al. 2011 and the references therein). TASEP enjoyed this attention since it is perhaps the simplest example of non-equilibrium statistical mechanics and its equilibrium distribution can be found analytically (Kolomeisky 1998). In the last 15 years this model has been

applied to biological systems including the translation process (Tripathy and Barma 1998, Harris and Stinchcombe 2004, Shaw et al. 2003, Zia et al. 2011). The TASEP model consists of particles that stochastically advance along a one-dimensional collection of discrete sites, in a preferred direction and under the exclusion constraint that no two particles can occupy the same position. It is important to note that in the translation process the advancement rates do depend on the position, but not on time. This is a consequence of tRNA availability which may be codon dependent, but this availability does not change appreciably on the time scale of a translation event. The randomness in advancement rate which is only space dependent is referred to as quenched randomness. A stationary distribution of such a process is well defined and has been studied by the TASEP community (Tripathy and Barma 1998, Harris and Stinchcombe 2004, Klumpp and Lipowsky 2004, Shaw et al. 2003, Zia et al. 2011).

This paper moves beyond the stationary distribution description and seeks to model time-dependent behavior of the bio-polymerization processes. Within the transcription process, the pauses change the advancement propensities in both time and space; therefore, our model is well-suited to describe the transcription process under these conditions.

The second class of elongation models are ordinary differential equation (ODE) models. The pioneering work in this area had been done by MacDonald, Gibbs and Pipkin (MacDonald et al. 1968, MacDonald and Gibbs 1969). Their model consists of n ODEs where n is the length of the template strand. The state variables x_i are the probabilities that the i -th position is occupied by the front of the elongating machine. This description allows modeling of spatially extended particles, which are appropriate for both transcription and translation processes.

The third class of models are those where both time and the template strand is assumed to be continuous. In the physics community continuum models have been used to gain insight into stationary distributions of TASEP for many years (Zia et al. 2011). Under simplified assumptions of a constant elongation speed and a termination speed equal to that of elongation speed, Heinrich and Rapoport (1980) introduced a simple delay model for initiation rate $\eta(t)$ and the concentration of the bound ribosomes $r_B(t)$. A recent paper by Mier-y-Teran-R et al. (2010) revisits and considerably improves the Heinrich and Rapoport model. Starting with the same system of ODEs for occupancy probabilities they explicitly derive a linear hyperbolic partial differential equation model, and from it a delay equation. Their main assumption is that the elongation rates change very little in both time and space.

The primary mathematical goal of this paper is to show that the classical nonlinear traffic model introduced in the 50's (Lighthill and Whitham 1955) can be rigorously linked to the ordinary differential models (MacDonald et al. 1968, MacDonald and Gibbs 1969, Heinrich and Rapoport 1980, Mier-y-Teran-R et al. 2010) for the motion of the polymerase on DNA. The nonlinear PDE model can be used to investigate transcription at high polymerase density, and in the presence of pauses, both of which lead to development of traffic jams.

In the traffic flow literature (for a recent review, Bellamo and Dogbe 2011), the macroscopic models examine the overall (average) behavior of the system. Variables considered in such models are the density of cars on the roadway and the flow of traffic over a prescribed length of roadway, see Rascle 2002. It is assumed that these models provide a good approximation when there are many cars on a single lane roadway, when the roadway is very long and when the measurements are taken over a long time scale (Gazis et al. 1961). Motivated by work undertaken in Newell (1961), Aw et al. (2002) establish a rigorous relationship between a PDE model and a *follow-the-leader* microscopic model consisting of a system of first order ODEs. The arguments lead to a comparison of discretizations of the microscopic and macroscopic models, and the PDE model is shown to be the limit of the time discretized ODE model, with appropriate variable transformations and scaling issues in both time and space resolved. That discussion, as well as others of a similar flavor of model development and analysis, see Argall et al. (2002), Daganzo (1995) and Aw and Rascle (2000), is given in the context of a traffic flow model.

In the spirit of Aw et al. (2002), we show that the difference equation formed by a time discretization of a particular ODE model is identical to the equation formed by the finite volume numerical method solving a nonlinear hyperbolic partial differential equation (PDE). Using this observation we describe how the solution of the ODE converges to the set of weak solutions of the PDE.

Before we apply the PDE traffic flow model to transcription of the ribosomal *rrn* operon, we compare the solution of the PDE to the solutions of the original continuous time Markov process in the presence of a single pause. We note that we do not expect a perfect agreement between these two solutions. The discrepancy is related to a fundamental problem of a microscopic structure of macroscopic shocks, that has been studied vigorously in the statistical physics community (Wick 1985, Ferrari et al. 1991, Derrida et al. 1993, Derrida et al. 1997). It has been shown that in an asymmetric exclusion processes, of which TASEP is a special case, and starting from an initial condition where density is piecewise constant with a unique shock, there exists a stationary continuous density profile which bridges the two initial densities as the spatial variable converges to $\pm\infty$ (Derrida et al. 1997). This behavior is absent in the PDE approximation. Since we are interested in finite time behavior on a finite spatial domain, the stationary density estimates cannot be used to estimate the discrepancy. However, our numerical simulation of biologically relevant examples show that the PDE based estimate of the induced delay is about 85% of the stochastic model estimate.

Finally, we apply our model to the ribosomal RNA operon. Note that elongation speeds have been observed experimentally in cells with mutated operons (Condon et al. 1993). In that setting, an average crossing time of approximately 60 seconds is measured for a strand of length 5450 nucleotides, and a corresponding elongation rate of 91 nt/s is estimated. This is based on the assumption that the velocity of an individual RNAP is approximately constant during the transcription process, not taking into account the poly-

merase pause mechanism that is known to occur. This estimate also assumes that the elongation rate is unaffected by the density of polymerases on the strand. We use the assumptions and analysis of the nonlinear PDE model to refine that estimate of the elongation rate under more realistic biological assumptions. First, assuming there are no polymerase pauses and using an experimentally reported estimate of the density of polymerases, we show that in order to achieve an observed crossing time of 60 seconds, the elongation rate of the individual polymerases must be approximately 132 nt/s. The difference between our estimate and that of the original estimate of 91nt/s by Condon et al. (1993) is attributed to the crowding effects of the polymerases that is accounted for in the PDE model. If we then assume that pauses in the rrn operon occur with the same frequency and are of the same duration as those in regular genes, we find that the total crossing time of a polymerase stays over 4 minutes for a range of biologically realistic values of elongation rates of individual polymerases. Since this is significantly larger than the observed average crossing time of 60 s, we conclude that an anti-termination complex, which is known to help rescue polymerases from termination sites (Albrechtsen et al. 1990, Dennis et al. 2009, Klumpp and Hwa 2008), must also assist in shortening of ubiquitous short pauses. Our results indirectly support conclusions of Klumpp and Hwa (2008) who have shown, using a stochastic model, that even a few non antiterminated paused polymerases can significantly reduce the average transcription rate. We hope that our model will be used to answer other important questions concerning the biology of cellular polymerization processes.

2 Background

Since we are not interested in the biological mechanism of RNAP pausing, we will not consider detailed stochastic models that take into account DNA information. Our starting point will be a TASEP model where DNA is a 1-D homogenous ribbon.

2.1 TASEP Process

A Totally Asymmetric Simple Exclusion Process (TASEP) is a lattice model of particles hopping stochastically along a 1-dimensional discrete lattice. This model was introduced in Spitzer 1970, and studied in various contexts (Janowsky and Lebowitz 1992, Derrida et al. 1993, Schutz and Domany 1993, Barkema et al. 1994, Kandel and Mukamel 1992, Kolomeisky 1998, Zia et al. 2011, Chou et al. 2011). We consider a TASEP with open boundaries (Kolomeisky 1998), whose exact stationary solutions are known (Derrida et al. 1993, Schutz and Domany 1993). Each RNAP is represented as a particle that can occupy exactly one position along the DNA. While the spatial extent of the polymerase has been studied from the modeling standpoint in the 1960s (MacDonald et al.

1968, MacDonald and Gibbs 1969) and may change the shape of the flux function (Zia et al. 2011), in this paper we concentrate on a particle of length corresponding to one nucleotide. There are three continuous time Markov processes that characterize the system. At the first position along the chain, a free RNAP has a probability αdt to start a new transcription, provided this position is empty. At all other positions the RNAP has probability βdt to advance to the next position, provided that it is empty, and at the last position RNAP moves off of DNA with probability γdt . The exact equilibrium solution depends on the relative size of the parameters α, β and γ (Kolomeisky 1998). When $\alpha < \gamma$ and $\alpha < \beta/2$ we are in the *low density phase* with equilibrium flux $\beta\alpha(1 - \alpha)$; when $\alpha > \gamma$ and $\gamma < \beta/2$ we are in *high density phase* with equilibrium flux $\beta\gamma(1 - \gamma)$; and finally, when both $\alpha > \beta/2$ and $\gamma > \beta/2$ we are in the *maximal flux case* with flux $\beta/4$. These cases differ by which rate dominates the motion: in the first case it is inflow rate α , second case outflow rate γ and in the third case the elongation rate β . The literature on TASEP models is quite extensive. The best recent reviews of this work are Chou et al. (2011) and Zia et al. (2011).

2.2 ODE models

A classical model of MacDonald and Gibbs (MacDonald et al. 1968, MacDonald and Gibbs 1969) models mean occupancy at each position of the strand and therefore consists of k ordinary differential equations where k is the length of the template strand (see also Heinrich and Rapoport 1980). The state variables z_j are the probabilities that the j -th position is occupied by the front of the elongating machine. In terms of the underlying stochastic process $z_j(t)$ can be thought of as a mean of the ensemble of random realizations of the stochastic process at a position j and time t . The state of z_j is determined by the balance of the rate of elongation v_j of the front from position j to $j + 1$ and the rate of elongation z_{j-1} from position $j - 1$ forward to j

$$\frac{dz_j}{dt} = v_{j-1} - v_j. \quad (1)$$

The elongation rate for *most* of the codons takes the form of a nonlinear expression that accounts for conditional probability that the $i + 1$ st codon is empty given that the i th codon is occupied. If we assume that the polymerase occupies one position along the strand, the elongation rate v_j takes the form

$$v_j := \beta_j z_j (1 - z_{j+1}). \quad (2)$$

Here β_j is the *rate* at which a polymerase moves between the position j and $j + 1$, z_j is the probability that the position j is occupied and $(1 - z_{j+1})$ is the probability that the position $j + 1$ is empty. There are two special equations, which model the influx of the polymerases at position 1 and the efflux at the last position k . For the influx equation we will assume that there is a large

reservoir of available polymerases and therefore the equation for z_1 includes the initiation rate

$$v_0 := \alpha r. \quad (3)$$

If the polymerase pool is limited (Mier-y-Teran-R et al. 2010), then

$$v_0 := \alpha \left(r - \sum_{i=1}^k z_i \right) (1 - z_1)$$

where α is the initiation rate, r is the total pool of polymerases and therefore $r - \sum_{i=1}^k z_i$ are the polymerases available for binding the DNA. Finally, the equation for z_k has the term $v_k = \gamma z_k$ where γ is the termination rate.

2.3 Continuum models

Mier-y-Teran-R et al. (2010) use an ODE model similar to (1) to derive a hyperbolic linear partial differential equation model

$$\partial_t(z(s, t)) + \partial_s(\beta(s, t)z(s, t)) = 0 \quad \text{for } 0 < s < L, t > 0 \quad (4)$$

with a specified initial condition $z(s, 0)$ and a nonlinear boundary condition $z(0, t)$, which reflects the global availability of polymerases for transcription initiation. This model replaces discrete binding sites along the template strand by a continuous position variable $s \in [0, n]$, where n is the number of binding sites in the discrete model of the template strand. The key parameter is the function $\beta(s, t)$ which characterizes the speed at which RNAP passes through position s along the template at time t : the higher the β , the greater the RNAP speed. The main assumptions necessary to derive a linear PDE are that the parameters of the system, including the $\beta(s, t)$, vary on a slower time scale than the time of RNAP transition between neighboring sites; that $\beta(s, t)$ varies slowly with s ; and that the distribution of the RNAP along the template is slowly varying.

Since we are interested in studying the effect of pauses on the transcription dynamics, we do not wish to impose the assumption that $\beta(s, t)$ slowly varies with s . Motivated by the form of the flux in the ODE model (2) we consider a nonlinear macroscopic traffic flow model

$$\partial_t z + \partial_x [\beta(x, t) z(1 - z)] = 0, \quad 0 < x < L, t > 0, \quad (5a)$$

$$z(x, 0) = z_0(x), \quad 0 \leq x \leq L \quad (5b)$$

Similarly to the ODE model, the influx boundary condition may reflect total availability of the polymerase

$$z(0, t) = \frac{\alpha}{\beta(0, t)} \left(r - \int_0^L z(x, t) dx \right) \quad t > 0, \quad (5c)$$

or assume large available pool of polymerases

$$z(0, t) = \frac{\alpha}{\beta(0, t)} r.$$

The function $z(x, t)$ represents a density, and we define

$$f(z) = \beta z (1 - z), \quad (6)$$

which is referred to as the flux function. The solution of equations (5), $z(x, t)$, represents a density of polymerases in space and time. As the choice of flux function in (6) is motivated by the ODE model where spatial extent of the RNAP is assumed to be one nucleotide, it should be noted here that if one considers the case where the spatial extent covers several nucleotides, then the function f will have a more complicated algebraic expression. This has been studied using the nomenclature “extended particles” in the TASEP literature, but we do not address the issue as it applies to the PDE model in the current work.

3 Discretization of ODE and PDE models

In this section we describe how the time discretization of the ODE system (1) can be used to design a convergent finite volume approximation for the PDE model (5). In the ODE model, equation (3), as well as the PDE model, (5c), the initiation rate is described in terms of an assumed total pool of polymerases available for initiation. Since the focus of this paper is to quantify effects of RNAP pausing on the transcription process, we assume that the initiation rate is constant for the following derivation (Kolomeisky 1998, Zia et al. 2011, Chou et al. 2011).

3.1 The PDE Model and its Finite Volume Method Approximation

In this section, we apply an elementary Godunov scheme to the PDE model in (5). Techniques from the theory of finite volume methods are used to analyze the numerical scheme, and convergence in the appropriate sense is addressed (LeVeque 2002). In order to develop a finite volume scheme for (5) over a finite interval of time $[0, T]$, we discretize the domain $[0, L] \times [0, T]$ using a collection of space-time mesh cells defined as $(x_{i-1/2}, x_{i+1/2}) \times [t_n, t_{n+1})$, for $i = 1, 2, \dots, K$ and $n = 0, 1, 2, \dots, N$, with $\Delta x = x_{i+1/2} - x_{i-1/2}$ and $\Delta t = t_{n+1} - t_n$. Denote the i th grid cell by $C_i = (x_{i-1/2}, x_{i+1/2})$, and recall that the integral form of the conservation law over C_i is

$$\frac{d}{dt} \int_{C_i} z(x, t) dx = f(z(x_{i-1/2}, t)) - f(z(x_{i+1/2}, t)).$$

Integrating this expression from t_n to t_{n+1} , dividing by Δx and rearranging terms leads to

$$\begin{aligned} \frac{1}{\Delta x} \int_{C_i} z(x, t_{n+1}) dx &= \frac{1}{\Delta x} \int_{C_i} z(x, t_n) dx \\ &\quad - \frac{1}{\Delta x} \left[\int_{t_n}^{t_{n+1}} f(z(x_{i+1/2}, t)) dt - \int_{t_n}^{t_{n+1}} f(z(x_{i-1/2}, t)) dt \right] \end{aligned}$$

Finite Volume Methods (FVMs) seek to design a numerical method of the form

$$Z_i^{n+1} = Z_i^n - \frac{\Delta t}{\Delta x} \left[F_{i+1/2}^n - F_{i-1/2}^n \right], \quad (7)$$

where the quantities Z_i^n approximate the averages $\frac{1}{\Delta x} \int_{C_i} z(x, t_n) dx$. Notationally, we associate a grid function with the discrete values Z_i^n and define the piecewise constant function $Z^{(\Delta t)}(x, t)$ which takes on the values

$$Z^{(\Delta t)}(x, t) = Z_i^n \quad \text{for } (x, t) \in [x_{i-1/2}, x_{i+1/2}) \times [t_n, t_{n+1}). \quad (8)$$

Methods are constructed so that the quantities Δt and Δx satisfy some suitable fixed relationship for all calculations. This relationship will be given for the specific problem considered here in a later section. FVMs also aim to design the flux function $F_{i-1/2}^n$ in such a way that the grid function converges in L^1 to a weak solution of the PDE on any finite interval of time.

Applying these ideas to the traffic flow model in (5), we construct a discrete set of coefficients $\hat{\beta}_i^n$ which represent average values of the original function $\beta(x, t)$ over each grid cell

$$\hat{\beta}_i^n := \frac{1}{\Delta x} \int_{C_i} \beta(x, t_n) dx, \quad (9)$$

and the flux function

$$F_{i-1/2}^n(Z_{i-1}^n, Z_i^n) = \hat{\beta}_{i-1}^n Z_{i-1}^n (1 - Z_i^n) \quad (10)$$

for all i .

With these choices the finite volume approximation equation (7) takes the form

$$Z_i^{n+1} = Z_i^n - \frac{\Delta t}{\Delta x} \left[\hat{\beta}_i^n Z_i^n (1 - Z_{i+1}^n) - \hat{\beta}_{i-1}^n Z_{i-1}^n (1 - Z_i^n) \right], \quad (11)$$

and under appropriate conditions, the theory shows that Z_i^n is an accurate approximation to the average value of $z(x, t)$ in the sense that

$$Z_i^n \approx \frac{1}{\Delta x} \int_{C_i} z(x, t_n) dx, \quad (12)$$

for all relevant i and n . Prior to the analysis of the scheme defined above, we first show that this scheme is identical with the time discretization of the ODEs in (1)-(2).

3.2 Discretization of the ODE

We consider a numerical discretization of (1)-(2) using an Euler scheme. The time domain is partitioned into a discrete set of equally spaced nodes $0 = t_0 < t_1 < \dots < t_N$ with $t_i - t_{i-1} = \Delta t$ for all $i = 1, 2, \dots, N$. Using the notation $z_j^n := z_j(t^n)$, we obtain a discrete system of the form

$$z_j^{n+1} = z_j^n - \Delta t \left[\beta_j(t_n) z_j^n (1 - z_{j+1}^n) - \beta_{j-1}(t_n) z_{j-1}^n (1 - z_j^n) \right]. \quad (13)$$

Therefore if we identify

$$Z_j^n = z_j^n \quad \text{and} \quad \hat{\beta}_j^n \Delta x = \beta_j(t_n) \quad (14)$$

then the discretization of the ODE equation (13) has the same form as finite volume approximation of the PDE (11).

By definition of $\hat{\beta}_j^n$ in (9) we observe that the rates $\{\beta_j(t)\}_{j=1}^k$ for the ODE are spatial accumulation (or the magnitude) of the continuous elongation rate $\beta(x, t)$ over the corresponding grid cell C_i

$$\beta_j(t_n) = \int_{C_j} \beta(x, t_n) dx = \int_{x_{j-1/2}}^{x_{j+1/2}} \beta(x, t_n) dx. \quad (15)$$

Now we are ready to state the main result.

4 Main result

For the statement of the following theorem, the notation $Z^{(\Delta t)}(x, t)$ represents the grid function introduced in the previous section: it is a piecewise constant function that takes the value Z_i^n on the space-time mesh cell $(x_{i-1/2}, x_{i+1/2}) \times [t_n, t_{n+1}) \subset [0, L] \times [0, T]$. When discussing convergence of a numerical method for a nonlinear conservation law, we must remember that a weak solution to the PDE may not be unique; hence, we define the set \mathcal{W} to be the set of all weak solutions,

$$\mathcal{W} = \{z : z(x, t) \text{ is a weak solution to (5)}\}$$

Then the global error for a given grid function, $Z^{(\Delta t)}$, is defined by

$$\text{dist}_T(Z^{(\Delta t)}, \mathcal{W}) = \inf_{z \in \mathcal{W}} \|Z^{(\Delta t)} - z\|_{1,T}.$$

The convergence result in this paper is proved in the sense that $\text{dist}(Z^{(\Delta t)}, \mathcal{W}) \rightarrow 0$ as the mesh size goes to zero.

More precisely, we fix $T > 0$ for any given sequence of mesh sizes $\Delta x \rightarrow 0$ we chose a corresponding sequence of $\Delta t \rightarrow 0$. For this choice of grid we define a grid function $Z^{(\Delta t)}(x, t)$ and denote by $\Delta t \rightarrow 0$ the fact that the grid size in both x and t converges to zero in this fashion.

Theorem 41 *We fix a domain $[0, L] \times [0, T]$, assume that the function $\beta(x, t)$ has finite total variation and let $B := \|\beta(x, t)\|_{1, T}$. For any Δx , let $\Delta t < \Delta t_0$, where*

$$\Delta t_0 = \min\left\{2\Delta x, \frac{\Delta x}{2B}\right\}.$$

Let $Z^{(\Delta t)}(x, t)$ be a grid function defined by (8). Then

$$\text{dist}_T(Z^{\Delta t}, \mathcal{W}) \rightarrow 0 \quad \text{as } \Delta t \rightarrow 0$$

where \mathcal{W} is a set of weak solutions of the hyperbolic PDE traffic model (5).

Since each grid function encodes a time discretization of ODE, this shows the relationship between the solutions of the discretized version of the PDE model (5) and that of the ODE model (1)-(2).

4.1 The main result

In this section we prove our main result, and we begin with some terminology that is essential for the analysis that follows; the most relevant definition is that of a *total variation* of a function. For an arbitrary function $Z(x)$, the total variation of Z can be defined as

$$TV(Z) = \sup \sum_{j=1}^N |Z(\xi_j) - Z(\xi_{j-1})|,$$

where the supremum is taken over all possible subdivisions of the real line $-\infty = \xi_0 < \xi_1 < \dots < \xi_N = \infty$. For the particular case where Z is assumed to be a grid function, one can define the total variation of Z as

$$TV(Z) = \sum_{i=-\infty}^{\infty} |Z_i - Z_{i-1}|,$$

and we note that $Z(x)$ must approach constant values z^\pm as $x \rightarrow \pm\infty$ in order for $TV(Z)$ to be finite.

Definition 42 *A numerical method is total-variation-stable, or TV-stable, if all the approximations $Z^{(\Delta t)}$ for $\Delta t < \Delta t_0$ lie in some fixed, compact set.*

The definition given above does not explicitly describe the form of the compact set; however, the following theorem from LeVeque (2002) shows that a uniform bound on the total variation is sufficient for a containment in the aforementioned compact subset of L^1 .

Theorem 43 *Consider a conservative method with a Lipschitz-continuous numerical flux and suppose that for each initial data Z^0 , there exist some $\Delta t_0, R > 0$ such that*

$$TV(Z^{\Delta t}) \leq R \quad \forall n, \Delta t \text{ with } \Delta t < \Delta t_0, \quad n\Delta t \leq T.$$

Then the numerical method is TV-stable.

With the previous machinery at hand, we will use the following convergence result to prove the main result (41)

Theorem 44 (LeVeque 2002) *Suppose $Z^{(\Delta t)}$ is generated by a numerical method in conservation form with a Lipschitz-continuous numerical flux, consistent with some scalar conservation law. If the numerical method is TV-stable, then the method is convergent in the sense that $\text{dist}_T(Z^{\Delta t}, \mathcal{W}) \rightarrow 0$ as $\Delta t \rightarrow 0$.*

4.2 Outline of the proof of Theorem 41

With the definition of the flux function as given in equation (10), the expression in (7) represents a discretization that is said to be in conservation form LeVeque (2002). Therefore, in view of Theorem 44 and Theorem 43, to prove Theorem 41 it is sufficient to establish a uniform bound on $TV(Z^{\Delta t})$ as $\Delta t \rightarrow 0$. This is the content of the following Theorem.

Theorem 45 *Assume $\beta(x, t)$ is a function of bounded total variation. Then for any $0 \leq t \leq T$ we have*

$$TV(Z^{(\Delta t)})(t) \leq R,$$

where the constant R is independent of Δt . In other words, the total variation is uniformly bounded as $\Delta t \rightarrow 0$.

4.3 Proof of Theorem 45

Before we start the estimates of the total variation we will need the following technical result.

Lemma 46 *Assume that a collection of initial values for the grid function $\{Z_i^0\}$ satisfies the inequality*

$$0 \leq Z_i^0 \leq 1.$$

Then if $\Delta t < \frac{\Delta x}{2B}$, the collection $0 \leq Z_i^n \leq 1$ for all i and n . That is, the numerical approximation of density, generated via (11), is bounded $0 \leq Z_i^n \leq 1$ for all relevant i, n .

Proof We proceed by induction. Assume that $0 \leq Z_{i-1}^n, Z_i^n, Z_{i+1}^n \leq 1$. We will show that with the choice of Δt above, we have $0 \leq Z_i^{n+1} \leq 1$. We write

$$\begin{aligned} Z_i^{n+1} &= Z_i^n - \frac{\Delta t}{\Delta x} [\beta_i Z_i^n (1 - Z_{i+1}^n) - \beta_{i-1} Z_{i-1}^n (1 - Z_i^n)] \\ &= Z_i^n - \frac{\Delta t}{\Delta x} \beta_i Z_i^n (1 - Z_{i+1}^n) + \frac{\Delta t}{\Delta x} \beta_{i-1} Z_{i-1}^n (1 - Z_i^n) \end{aligned}$$

If $Z_i^n = 1$ then the last equation reads

$$Z_i^{n+1} = 1 - \frac{\Delta t}{\Delta x} \beta_i (1 - Z_{i+1}^n) \leq 1 - \frac{\Delta t}{\Delta x} \beta_i < 1.$$

Now we assume that $Z_i^n = 1 - \epsilon$. Then the equation above reads

$$\begin{aligned} Z_i^{n+1} &= 1 - \epsilon - \frac{\Delta t}{\Delta x} \beta_i (1 - \epsilon) (1 - Z_{i+1}^n) + \frac{\Delta t}{\Delta x} \beta_{i-1} Z_{i-1}^n \epsilon \\ &\leq 1 - \epsilon + \frac{\Delta t}{\Delta x} \beta_{i-1} Z_{i-1}^n \epsilon \\ &\leq 1 - \epsilon + \epsilon \frac{\Delta t}{\Delta x} B. \end{aligned}$$

With the choice of time step $\Delta t < \frac{\Delta x}{2B}$ we get $Z_i^{n+1} < 1$. The proof of $Z_i^{n+1} \geq 0$ proceeds similarly.

Now we begin to estimate the total variation of the grid function $Z^{(\Delta t)}$. We rewrite equation (7) as

$$\begin{aligned} Z_i^{n+1} &= Z_i^n - \frac{\Delta t}{\Delta x} (F_{i+1/2} - F_{i-i/2}) \\ &= Z_i^n - \frac{\Delta t}{\Delta x} (F_{i+1/2} - f(Z_i) + f(Z_i) - F_{i-i/2}) \\ &= Z_i^n - \frac{\Delta t}{\Delta x} \frac{f(Z_i^n) - F_{i-1/2}}{Z_i^n - Z_{i-1}^n} (Z_i^n - Z_{i-1}^n) + \frac{\Delta t}{\Delta x} \frac{f(Z_i^n) - F_{i+1/2}}{Z_{i+1}^n - Z_i^n} (Z_{i+1}^n - Z_i^n) \\ &=: Z_i^n - C_{i-1}^n (Z_i^n - Z_{i-1}^n) + D_i^n (Z_{i+1}^n - Z_i^n) \end{aligned} \quad (16)$$

where

$$C_{i-1}^n := \frac{\Delta t}{\Delta x} \frac{f(Z_i^n) - F_{i-1/2}}{Z_i^n - Z_{i-1}^n}, \quad D_i^n := \frac{\Delta t}{\Delta x} \frac{f(Z_i^n) - F_{i+1/2}}{Z_{i+1}^n - Z_i^n}.$$

Our estimate below follows closely the argument in Harten's Theorem (LeVeque 2002). However, that result is not directly applicable since the conductivity β_i^n in the flux function F may vary in both i and n . We use the form (16) and estimate the total variation of this scheme by carefully bounding the terms C_{i-1}^n and D_i^n . To obtain an estimate for D_i^n , we add and subtract the term $\beta_i Z_i^n (1 - Z_i^n)$ between the first and second lines below

$$\begin{aligned} D_i^n &= \frac{\Delta t}{\Delta x} \frac{\beta(x_i) Z_i^n (1 - Z_i^n) - \beta_i Z_i^n (1 - Z_{i+1}^n)}{Z_{i+1}^n - Z_i^n} \\ &= \frac{\Delta t}{\Delta x} \left[\beta_i \frac{Z_i^n (1 - Z_i^n) - Z_i^n (1 - Z_{i+1}^n)}{Z_{i+1}^n - Z_i^n} + \frac{\beta(x_i) - \beta_i}{Z_{i+1}^n - Z_i^n} Z_i^n (1 - Z_i^n) \right] \\ &= \frac{\Delta t}{\Delta x} \left[\beta_i Z_i^n + \frac{\beta(x_i) - \beta_i}{Z_{i+1}^n - Z_i^n} Z_i^n (1 - Z_i^n) \right] \\ &= \bar{D}_i^n + Q_i^n, \end{aligned}$$

where by Lemma 46

$$\bar{D}_i^n := \frac{\Delta t}{\Delta x} \beta_i Z_i^n \geq 0 \quad \text{and} \quad Q_i^n := \frac{\Delta t}{\Delta x} \frac{\beta(x_i) - \beta_i}{Z_{i+1}^n - Z_i^n} Z_i^n (1 - Z_i^n).$$

Estimating C_{i-1}^n , we will have an extra term after adding and subtracting the term $\beta_{i-1}Z_i^n(1-Z_i^n)$ between the first and second lines below

$$\begin{aligned} C_{i-1}^n &= \frac{\Delta t}{\Delta x} \frac{\beta(x_i)Z_i^n(1-Z_i^n) - \beta_{i-1}Z_{i-1}^n(1-Z_i^n)}{Z_i^n - Z_{i-1}^n} \\ &= \frac{\Delta t}{\Delta x} [\beta_{i-1} \frac{Z_i^n(1-Z_i^n) - Z_{i-1}^n(1-Z_i^n)}{Z_i^n - Z_{i-1}^n} + \frac{\beta(x_i) - \beta_{i-1}}{Z_i^n - Z_{i-1}^n} Z_i^n(1-Z_i^n)] \\ &= \frac{\Delta t}{\Delta x} [\beta_{i-1}(1-Z_i^n) + \frac{\beta(x_i) - \beta_{i-1}}{Z_i^n - Z_{i-1}^n} Z_i^n(1-Z_i^n)] \\ &= \bar{C}_{i-1}^n + R_{i-1}^n, \end{aligned}$$

where by Lemma 46

$$\bar{C}_{i-1}^n := \frac{\Delta t}{\Delta x} \beta_{i-1}(1-Z_i^n) \geq 0 \quad \text{and} \quad R_{i-1}^n := \frac{\Delta t}{\Delta x} \frac{\beta(x_i) - \beta_{i-1}}{Z_i^n - Z_{i-1}^n} Z_i^n(1-Z_i^n).$$

Note that both $\bar{D}_i^n \leq 1$ and $\bar{C}_i^n \leq 1$ when $\Delta t < \frac{\Delta x}{B}$. Furthermore, since $\Delta t < \frac{\Delta x}{2B}$ we have

$$\bar{D}_i^n + \bar{C}_i^n = \frac{\Delta t}{\Delta x} \beta_{i+1}[1 - Z_{i+1}^n + Z_i^n] < 1.$$

Using (16) and the expressions for C_i^n and D_i^n we estimate the total variation at step $n+1$ as follows:

$$\begin{aligned} TV(Z^{n+1}) &= \sum_i |Z_{i+1}^{n+1} - Z_i^{n+1}| \\ &= \sum_i |(Z_{i+1}^n - Z_i^n)(1 - \bar{C}_i^n - \bar{D}_i^n - R_i^n - Q_i^n) + (\bar{D}_{i+1}^n + Q_{i+1}^n)(Z_{i+2}^n - Z_{i+1}^n) \\ &\quad + (\bar{C}_{i-1}^n + R_{i-1}^n)(Z_i^n - Z_{i-1}^n)| \\ &\leq \sum_i |Z_{i+1}^n - Z_i^n| |1 - \bar{C}_i^n - \bar{D}_i^n| + |\bar{D}_{i+1}^n| |Z_{i+2}^n - Z_{i+1}^n| + |\bar{C}_{i-1}^n| |Z_i^n - Z_{i-1}^n| \\ &\quad + \sum_i |(Z_{i+1}^n - Z_i^n)(-R_i^n - Q_i^n) + Q_{i+1}^n(Z_{i+2}^n - Z_{i+1}^n) + R_{i-1}^n(Z_i^n - Z_{i-1}^n)| \\ &\leq \sum_i [|Z_{i+1}^n - Z_i^n| (1 - \bar{C}_i^n - \bar{D}_i^n) + \bar{D}_{i+1}^n |Z_{i+2}^n - Z_{i+1}^n| + \bar{C}_{i-1}^n |Z_i^n - Z_{i-1}^n|] \\ &\quad + \frac{\Delta t}{\Delta x} \sum_i |(\beta_i - \beta(x_{i+1}))Z_{i+1}^n(1 - Z_{i+1}^n) + (\beta_i - \beta(x_i))Z_i^n(1 - Z_i^n) \\ &\quad + (\beta(x_{i+1}) - \beta_{i+1})Z_{i+1}^n(1 - Z_{i+1}^n) + (\beta(x_i) - \beta_{i-1})Z_i^n(1 - Z_i^n)| \\ &= \sum_i [|Z_{i+1}^n - Z_i^n| (1 - \bar{C}_i^n - \bar{D}_i^n) + \bar{D}_{i+1}^n |Z_{i+2}^n - Z_{i+1}^n| + \bar{C}_{i-1}^n |Z_i^n - Z_{i-1}^n|] \\ &\quad + \frac{\Delta t}{\Delta x} \sum_i |Z_{i+1}^n(1 - Z_{i+1}^n)[-(\beta_{i+1} - \beta_i)] + Z_i^n(1 - Z_i^n)[\beta_i - \beta_{i-1}]| \end{aligned}$$

$$\begin{aligned}
&\leq \sum_i |Z_{i+1}^n - Z_i^n| + 2 \frac{\Delta t}{\Delta x} \sum_i |Z_{i+1}^n (1 - Z_{i+1}^n)| |(\beta_{i+1} - \beta_i)| \\
&\leq TV(Z^n) + \frac{1}{2} \frac{\Delta t}{\Delta x} \sum_i |(\beta_{i+1} - \beta_i)|
\end{aligned}$$

where we handled the first term by shifting the counting index down by one in terms multiplying \bar{D}_{i+1}^n and up by one in terms multiplying \bar{C}_{i-1}^n . In the last equation we have used that $x(1-x) < 1/4$ for $0 \leq x \leq 1$.

We finish by relating the last sum to the total variation of the function $\beta(x, t)$. Recall from (15) that for a fixed time t there is the following relationship between the function $\beta(x, t)$ and the collection of functions $\{\beta_i(t)\}_{i=1}^{L/\Delta x}$ used in the ODE model

$$\beta_i(t) = \int_{x_{i-1/2}}^{x_{i+1/2}} \beta(s, t) ds.$$

By the mean value Theorem

$$\beta_i = \beta(\xi_i) \Delta x$$

where $x_{i-1/2} \leq \xi_i \leq x_{i+1/2}$ for all i . This yields

$$\frac{1}{\Delta x} \sum_i |\beta_{i+1} - \beta_i| \leq \sum_i |(\beta(\xi_{i+1}) - \beta(\xi_i))| \leq TV(\beta) \quad (17)$$

and therefore we have

$$TV(Z^{n+1}) \leq TV(Z^n) + \frac{\Delta t}{2} TV(\beta). \quad (18)$$

As we refine our grid with $\Delta t \rightarrow 0$ the grid function at the end of the time interval $Z^{(\Delta t)}(T) = Z^n$ where n is an integer related to the step size Δt by $n = \lfloor \frac{T}{\Delta t} \rfloor$. It follows that $(n+1)\Delta t \leq 2T$ for $\Delta t \leq T$. Using this and the estimate (18) recursively we get

$$TV(Z^{n+1}) \leq TV(Z^0) + (n+1)\Delta t \frac{TV(\beta)}{2} \leq TV(Z^0) + TV(\beta)T.$$

This uniform bound depends on the initial condition $Z^{(0)}$, the total variation of conductivity $\beta(x, t)$ contained within the flux function f and the final time T but is independent of Δt .

This completes the proof of Theorem 45. Since Theorem 45 provides the remaining key estimate in the proof of Theorem 41, this finishes the proof of the main result.

Remark 1 If we assume that β is constant with respect to both space and time a simplified argument can be used to prove a stronger result. In particular, in that case there exists Δt sufficiently small such that the time Δt -discretization of the ODE system in equation (1) satisfies Harten's Theorem and therefore

is a total variation diminishing (TVD) method for solving the PDE model (5). Furthermore, the discretization of the ODE in (1)-(2) is an E-scheme for solving the PDE model (5); therefore, it is convergent to an entropy-satisfying weak solution of the PDE.

5 Applications

The previous analysis shows that if one uses a standard Godunov scheme to generate a numerical approximation to a solution of the nonlinear PDE given in (5), the resulting discrete system of equations defined by (7) and (10) is exactly the same as those that one derives when applying a standard Euler method to the system of ODEs described in (1)-(2). Therefore, we invoke a numerical study of the behavior of solutions of the nonlinear conservation law as a means to analyzing the underlying biological questions related to the transcription process. An advantage of using this particular nonlinear PDE is that closed form solutions can be written for several model problems of interest; while the system of nonlinear ODEs in (1)-(2) cannot be solved analytically. Polymerase pauses at bottleneck nucleotide locations along the DNA strand can be modeled in a very simple manner using this PDE. In order to gain insight into the effects these pauses have on total throughput of polymerases, we first present a basic model problem that is used to construct a closed form analytical solution for the simplest case of one pause location. The analytical solutions and the numerical computations demonstrate excellent agreement for the simple model. Once the proof-of-concept is demonstrated for that model, a more realistic case is included and predictions are made based on the numerical results.

5.1 Comparison of the PDE and Stochastic Models

5.1.1 PDE Model with One Pause Location

In this section, we examine the PDE model given in (5) with a discontinuous function $\beta(x, t)$ that allows us to model transcription behavior along a single DNA strand where one of the nucleotides corresponds to a pause location. This equation takes the form

$$z_t + [\beta(x, t)(1 - z)]_x = 0 \quad x \in (-0.5, 0.5), \quad t > 0 \quad (19)$$

$$z(x, 0) = z_0 \quad (20)$$

$$z(-0.5, t) = z_0 \quad (21)$$

where the piecewise constant function

$$\beta(x, t) = \begin{cases} 0, & \text{if } x = 0 \text{ and } 0 < t < \tau \\ 1, & \text{otherwise} \end{cases}$$

represents a raw elongation velocity function which assumes a maximum value of 1 except at the site of the pause $x = 0$ where it assumes the value of 0 for a duration of time length determined by the parameter τ . During this time period, $\beta(x, t)$ is set to 0 and no elongation is allowed to take place at this spatial location. This formulation models the location of a bottleneck nucleotide in the middle of the DNA strand (at $x = 0$), beginning at time $t = 0$ and ending at time $t = \tau$. Comparing this model to (5) we note that we assume an unlimited reservoir of polymerases at density z_0 , which determines both the initial condition and the boundary condition on the left. Note that this PDE is well-posed with a boundary condition on the left end of the spatial domain at $x = -0.5$, and it does not require a boundary condition on the right. By choosing $\beta(0.5, t) = 1$, we are tacitly assuming that the termination rate is the same as the elongation rate at the interior nucleotides. In order to align the model more closely with the TASEP model or the ODE model, a different choice of termination rate can be implemented by assigning $\beta(0.5, t)$ at an appropriate level. In the PDE model, the time variable t represents the dimensionless time, and we assume that $z_0 < 0.5$ so that the initial density allows the background flow (flux) to be below its maximum possible value.

Using the method of characteristics, one can develop a closed-form solution to (19)-(21), see Haberman (1998) for example. For $0 < t < \tau$, there are three shocks emanating from $x = 0$. As we cross the spatial domain (from left to right), to the left of the bottleneck, the density jumps from z_0 to 1 as the polymerases stack up behind the paused polymerase. At the spatial location of the bottleneck, the density jumps from its maximum value of 1 to its minimum value of 0, and then, further to the right, the density jumps from 0 to z_0 as the tail end of the polymerases not stopped at $x = 0$ continue their motion across the DNA strand. Using these shocks, a closed-form representation of the solution for $0 \leq t < \tau$ is given by

$$z(x, t) = \begin{cases} z_0, & x < -z_0 t \\ 1, & -z_0 t < x < 0 \\ 0, & 0 < x < (1 - z_0)t \\ z_0, & (1 - z_0)t < x \end{cases}$$

This expression for $z(x, t)$ describes the behavior of polymerase density during the time that the pause occurs. One can also describe the solution for $t \geq \tau$. At $t = \tau$, the bottleneck nucleotide (at $x = 0$) releases the lead polymerase. At this point, the paused polymerases begin to elongate, and a rarefaction wave is created from the maximum density, $z = 1$, to the minimum density, $z = 0$. The entire mass of stopped polymerases will dissipate in time after the bottleneck release; the boundary of this region is a shock that passes each position at a particular time, after which the density abruptly drops to z_0 . In the following section, we extend the formulation of the problem and then give a graph of a typical solution for a particular set of parameters, see Figure 1.

5.1.2 PDE Model with One Pause in a Parameterized Interval of Time

One can also explicitly solve a more general problem, where the pause at the bottleneck site does not occur at the initial time $t = 0$ but at some parameterized time interval denoted by $\zeta < t < \xi$. As in the previous section, this is accomplished by using a piecewise definition of the coefficient function β . Hence we solve the PDE given by (19)-(21) where the piecewise constant function $\beta(x, t)$ is now taken to be

$$\beta(x, t) = \begin{cases} 0 & \text{for } x = 0 \text{ and } \zeta < t < \xi \\ 1 & \text{otherwise.} \end{cases} \quad (22)$$

The solution to this PDE can be derived using the method of characteristics. For brevity, we omit the details of this calculation. To describe the solution we define specific times t_1 and t_2 in the following way: t_1 is the time at which the line of paused polymerases has completely dissipated after the paused nucleotide has released and the density is returned to the uniform value of z_0 . Further, t_2 is the time required for the lead polymerase that is stopped at the paused nucleotide to *catch up with* those polymerases that were unaffected by the pause once it has ended; ie, the density at the front end returns to the uniform value of z_0 . For visualizing these solutions, we give an example with $dt = \xi - \zeta$, $t_1 = \frac{dt}{1-z_0}$ and $t_2 = \frac{dt}{z_0}$ and $z_0 = \frac{1}{5} = 0.2$ and $\zeta = 0$, and the closed-form solution to (19)-(21) using (22) is given by

$$z(x, t) = \begin{cases} 0.2 & x < -0.2t; \quad t < t_1 \\ 1 & -0.2t < x < 0; \quad t < \xi \\ 1 & -0.2t < x < -(t - \xi); \quad \xi < t < t_1 \\ \frac{1}{2} - \frac{x}{2(t - \xi)} & -(t - \xi) < x < 0; \quad \xi < t < t_1 \\ 0.2 & x < [1 - 2(0.2)](t - \xi) - 2 \left[\frac{0.2\xi(t - \xi)}{1 - 0.2} \right]^{1/2} (1 - 0.2); \quad t \geq t_1 \\ \frac{1}{2} - \frac{x}{2(t - \xi)} & [1 - 2(0.2)](t - \xi) - 2 \left[\frac{0.2\xi(t - \xi)}{1 - 0.2} \right]^{1/2} (1 - 0.2) < x < 0; \quad t \geq t_1 \\ 0 & 0 < x < (1 - 0.2)t; \quad t < \xi \\ \frac{1}{2} - \frac{x}{2(t - \xi)} & 0 < x < (t - \xi); \quad \xi < t < t_2 \\ 0 & (t - \xi) < x < (1 - 0.2)t; \quad \tau < t < t_2 \\ 0.2 & (1 - 0.2)t < x; \quad t < t_2 \\ \frac{1}{2} - \frac{x}{2(t - \xi)} & 0 < x < [1 - 2(0.2)](t - \xi) + 2(0.2) \left[\left(\frac{1}{0.2} - 1 \right) (t - \xi)\xi \right]^{1/2}; \quad t \geq t_2 \\ 0.2 & [1 - 2(0.2)](t - \xi) + 2(0.2) \left[\left(\frac{1}{0.2} - 1 \right) (t - \xi)\xi \right]^{1/2} < x; \quad t \geq t_2 \end{cases} \quad (23)$$

Figure 1 depicts a contour plot of the solution $z(x, t)$ with parameters chosen as in equation (23). One can clearly distinguish the time of the pause at $x = 0$ and its duration. One can also discern the “spreading out” of the RNAP traffic once the pause has ended. In addition, the termination site is represented by the vertical axis at $x = 0.5$, and one can see that the pause duration of 0.1 time

units results in downstream effects in the RNAP density at the termination site that are discernible for more than one full time unit (note the solution behavior at $x = 0.5$ between time values of approximately $t = 1.2$ and $t = 2.2$). In the sections that follow, we use this expression to validate numerical simulations of the PDE and to study the delays experienced by polymerases in the presence of a single pause nucleotide.

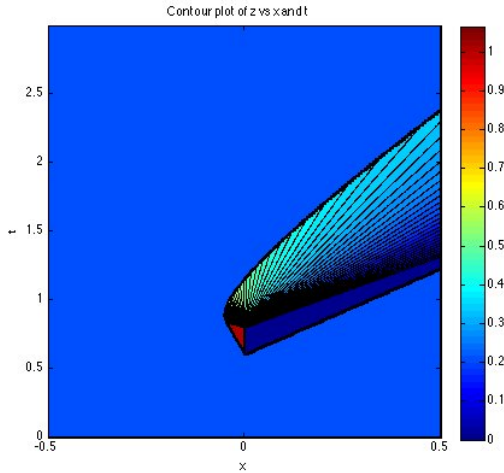


Fig. 1 A contour plot of the solution $z(x, t)$ with $z_0 = 0.2$ where $\beta(x, t)$ incorporates a pause at $x = 0$, $\zeta = 0.7$ and $\xi = 0.8$ from equation (22).

5.1.3 Stochastic Model with One Pause Location

This section describes an algorithm used to simulate the type of TASEP process described in Section 2.1. Stochastic simulation algorithms can be very slow, since they need to simulate a DNA strand thousands of nucleotides long (for the *rrn* gene it is 5450 nucleotides) with many polymerases on it at any given time. Therefore in our implementation we have focused on the efficiency of identifying the next event to be executed. There are two main constructs in our algorithm: the *array of positions* along the DNA strand and a *priority queue* for events. Each position in the array stores a set of pauses that occur at that position at some time. Each pause is encoded as a pair of numbers specifying the start and end times of the pause. An event has an associated action and a time stamp, which indicates when the action will happen.

The basic function of the program is to identify the event with the lowest time stamp and execute its function, which may include adding the event back into the queue. There are two basic types of events—*generator events* and *polymerase events*. The generator event corresponds to the start position (initiation location) of the DNA strand. The associated generator action is

to create new polymerase events. When a generator event occurs and the first position in the array is empty, the polymerase event is assigned the first position of the strand, and a time stamp is created. The time stamp encodes the time at which the polymerase will attempt to move to the next position. This time is computed in the following way. First, an exponentially distributed time τ with constant β is generated. If there is no pause at the current position, then τ is the time of the next move for the polymerase. If there is a pause, then the time τ is added to the time when the pause expires. After assigning the position and the time stamp, the polymerase event is added back onto the priority queue.

When the generator attempts to produce a polymerase event, it produces a time stamp which encodes the time at which the next polymerase will be created. This time stamp is created by producing an exponentially distributed time τ with constant α . After the time stamp is computed, the generator is then put back into the priority queue.

When the generator event comes to the top of the priority stack, it produces a polymerase event, a new time stamp and is put back into the queue. When the polymerase event comes to the top of the priority stack, it moves to the next position, computes its time stamp and is put back into the queue. When a polymerase reaches the right end (the last position in the array) of the DNA, it generates exponentially distributed exit time stamp with rate γ and subsequently records its data into a database. This data consists of the time of creation, the time of finish, and if there were intermediate checkpoints, the time when it passed these checkpoints. All density and flux information can be easily computed from this data. In our simulations of the *rrn* gene below we set $\gamma = \beta$.

5.2 Numerical Simulations for the PDE and Stochastic Models

5.2.1 Discontinuous Galerkin Scheme for PDE Model Simulations

Although Section 3.1 applies a Godunov type of scheme to the model in (5) in order to connect its full discretization with that of the system of ODEs described in (1), a Discontinuous Galerkin Finite Element Method (DG) is used to numerically approximate the solution of (19)-(22) and to numerically estimate delays discussed in a later section. DG is well-suited to the task of generating high order accuracy of solutions to a PDE with the discontinuity in the velocity coefficient $\beta(x, t)$ using a standard implementation. Prior to discussing the results of the simulations, we give a very brief overview of the DG approach. The interested reader is referred to a wealth of literature on the topic for more details, see Cockburn and Shu 2001, Hesthaven and Warburton 2008, Arnold et al. 2001/02 and the references contained therein.

Let the function z_h denote the DG approximation of the solution to equation (19)-(21) with (22). To obtain this approximation, one first discretizes the spatial domain, $[-0.5, 0.5]$, into \mathcal{K} elements, $D^k = [x_l^k, x_r^k]$ for $k = 1, 2, \dots, \mathcal{K}$.

On each element, choose a set of N interpolation points used to define a basis of Lagrange interpolating polynomials, $\{\ell_i^k(x)\}_{i=1}^N$. The semi-discrete DG formulation on the k^{th} element is

$$M^k \frac{d}{dt} \mathbf{z}_h^k + S^k \mathbf{f}_h^k = [\ell^k(x)(f_h^k - f^*)]_{x_l^k}^{x_r^k}$$

where the mass and stiffness matrices are

$$M_{ij}^k = \int_{x_l^k}^{x_r^k} \ell_j^k(x) \ell_i^k(x) dx, \quad S_{ij}^k = \int_{x_l^k}^{x_r^k} \ell_i^k(x) \frac{d\ell_j^k}{dx} dx$$

and f^* is a numerical flux defined at the interface used to connect the local solutions. Using a total variation diminishing Runge-Kutta method, this system of ODEs in time is solved to find the approximation $z_h(x, t)$, see Hesthaven and Warburton (2008) and references therein. Matlab code obtained from the website associated with Hesthaven and Warburton (2008) provided the foundational code used for the PDE simulations presented here. Note that similar calculations for a particular instance of the model in (19) using DG have appeared in print, see Zhang and Liu 2005. The results in this paper differ from those in several ways. In Zhang and Liu 2005, various combinations of two numerical flux functions and two flux limiters are applied for those calculations, and for each of those cases, performance of the numerical computations on a simple model is compared. In this paper, a nonlinear Lax-Friedrichs flux with a `minmod` slope limiter is used for all of the numerical simulations presented, and the current research effort focuses on the use of the numerical results obtained through DG in order to estimate delays for the particular biological application of polymerization.

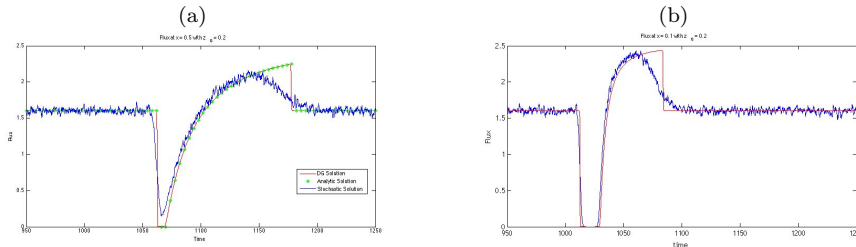


Fig. 2 (a) Three computations of density at the termination site of the DNA model using a background density of $z_0 = 0.2$. The blue graph represents the data from the stochastic TASEP model. The solid red curve is the graph of $z_h(0.5, t)$, the result of the DG simulation of the PDE model in (19). The dashed green curve represents the true solution as given in (23). (b) Comparison of the flux at position $x = 0.1$ using a background density of $z_0 = 0.2$. Blue and red curves are as in (a).

Figure 2 presents an example of the numerical calculations of density obtained using the DG calculations as well as the results of the simulation of the TASEP model. The initial and background density of $z_0 = 0.2$ is assumed for

all of the calculations, see equations (20) and (21). The density at the right boundary $x = 0.5$, representing the termination site, is plotted as a function of the time t . The dimensionless time variable used in (19)-(21) has been converted to dimensional time with units of seconds for the graphical comparison. For completeness, the true solution to the PDE, as it is given in equation (23) is plotted as well. In Figure 2, the DG calculation shows excellent agreement with the true solution of the PDE. Both the DG and the TASEP calculations exhibit the same overall trends, except within the region corresponding to roughly the time interval of 1150 to 1180 seconds.

We note that the observed discrepancy is related to a fundamental problem of a microscopic structure of macroscopic shocks, that has been studied vigorously in the statistical physics community (Wick 1985, Ferrari et al. 1991, Derrida et al. 1993, Derrida et al. 1997). The stationary density of the particles in asymmetric exclusion processes, of which TASEP is a special case, is described on suitable macroscopic spatial and temporal scales by the inviscid Burgers' equation (Andjel and Kipnis 1984, Andjel and Vares 1986, Wick 1985); the latter has shock solutions with a discontinuous jump. One of the main questions was whether the stationary density profile in the stochastic model exhibited the same discontinuity (Ferrari et al. 1991, Derrida et al. 1993, Derrida et al. 1997). It has been shown that, starting from an initial condition where density is piecewise constant with a unique shock (in other words, a Riemann problem), there exists a stationary continuous density profile which bridges the two initial densities as the spatial variable converges to $\pm\infty$ (Derrida et al. 1997). This profile exists only if one takes a view from the position of a particle which is initially inserted at the spatial location of the shock. This technical result and the corresponding recursive formulas for the stationary density cannot be used to estimate the discrepancy between the PDE and the stochastic exclusion process, since we are interested in the solutions at a finite time after the pause site. Furthermore, the initial condition after the pause has ended contains only a finite spatial interval at high density corresponding to polymerases stopped at (and backed up behind) the pause site.

With a series of numerical investigations, we have noted that the discrepancy between the TASEP and the PDE model is smaller when we measure flux at $x = 0.1$ rather than at $x = 0.5$, (that is, at a nucleotide location that is much closer to the pause location) see Figure 2b. Since the experimental data in the literature suggest that a pause is encountered on average every 100 nt, which in our scaling corresponds to a spatial scale of 0.1units, Figure 2b captures this discrepancy on a more realistic spatial scale.

5.3 Delay Computations for the Model Problems

Here we use our PDE model to quantify the effect that a realistic distribution of pauses has on the average time required for a polymerase to cross the DNA strand. We begin by presenting the ideas in the setting of a single pause, placed

in the middle of the DNA strand, as in the previous section. We also compare our results with the stochastic computation using the TASEP model. In view of Figure 2 we do not expect a perfect match in computed effect. Since the PDE simulation is more efficient and avoids conceptual issues described below, we seek an approximate fit and a quantifiable difference. We first outline our approach to quantification of the effect of the pause on polymerase traffic. We want to compare the average time required to cross the DNA strand with and without the pause, and since each polymerase is affected by a pause differently, we would like to compute a *pause-induced delay per affected polymerase*. However, there are some conceptual challenges with this notion.

In contrast to the PDE model where the density drops to the background density z_0 via a shock wave, in the stochastic model we cannot determine which polymerases are affected by the pause and which are not. Therefore we are forced to select arbitrary times T_0 and T_1 and compute the pause-induced delay for all polymerases that arrive at the end of the strand between these two times. We select these times so that they include the density disturbance seen in Figure 2. Since we may include some polymerases that were not affected by the pause, the delay per polymerase will depend on T_0 and T_1 . The situation is more straightforward for the PDE model where the affected polymerases arrive at times that are clearly delineated. We use the same times T_0 and T_1 for both PDE and stochastic model, and since our main goal is to ensure that the PDE computation does not stray too far from the stochastic model, we did not try to address the dependence on T_0 and T_1 by averaging over periodically occurring pauses.

5.3.1 Delay Computations Using the Stochastic Model

Let $T_a(i)$ denote the total time required for the i th polymerase to cross the DNA strand. Let μ denote the expected value of the time required for an arbitrary polymerase (that doesn't experience a pause) to cross the DNA strand. Choose T_0 to be a time after the ramp-up phase of the stochastic model has passed (so that the average density of polymerases on the DNA strand is approximately z_0) but prior to the time at which the pause location along the DNA has been activated. Similarly, choose T_1 to be a time after all of the polymerases affected by the pause location have reached the termination site. Define the index K_b to be the integer index corresponding to the first polymerase that arrives at the termination site at a time on or after T_0 . Similarly, the index K_e is the integer index corresponding to the last polymerase to arrive at the termination site at a time at or prior to T_1 . The average pause delay per affected polymerase is given by

$$D_S(T_0, T_1) = \frac{1}{K_e - K_b} \sum_{i=K_b}^{K_e} (T_a(i) - \mu) \quad (24)$$

5.3.2 Delay Computations Using the PDE Model

To compute the effect of the pause we compare the solution of (19)-(22) which incorporates one pause in the middle of the DNA with the of the same PDE model where the only change is that we set $\beta(x, t) = 1$ for all x and t . The second model characterizes the ideal situation in which there is no pause, and the initial and boundary conditions are specified with a certain constant *background* density of z_0 . Note that the constant density $z(x, t) = z_0$ is the solution of such an equation; therefore, the flow function, f_R , is also a constant function

$$f_R(x, t) = f_R = (1 - z_0) z_0$$

Since f_R is constant, then the number of polymerases that have arrived at the termination site, $x = 0.5$, in the time interval (T_0, t) is

$$N(t) = \int_{T_0}^t f_R dy = (t - T_0) f_R$$

The explicit solution to equation (19)-(22) has been constructed in the previous section. Denoting the flow function from (19) as $f(z) = \beta(x, t)(1 - z)z$, then for the system governed by this model, the number of polymerases that have crossed the termination site in the time measured from T_0 to the time t is given by

$$F(t) = \int_{T_0}^t f(z(0.5, y)) dy .$$

For any $t > T_0$, we seek to calculate the amount of time required for the polymerases that have been stopped by a pause to reach the right boundary and compare that time with the amount of time that is required for the same number of polymerases to arrive at the termination site under the condition that the DNA chain has no pause. To achieve this goal, we define the function $s(t)$ so that for each $t > T_0$, the function $s(t)$ is the instant of time for which

$$\int_{T_0}^{s(t)} f(z(0.5, y)) dy = N(t). \quad (25)$$

The average delay over an interval $[T_0, T_1]$ is then calculated to be

$$D_P(T_0, T_1) = \frac{1}{\int_{T_0}^{T_1} f(z(0.5, t)) dt} \int_{T_0}^{T_1} f(z(0.5, t))(s(t) - t) dt \quad (26)$$

where T_0 is an arbitrary time before the first of the delayed polymerases reached the right boundary and T_1 is an arbitrary time after the last of the polymerases affected by the pause has reached the termination site.

The previous equations assume that we have the analytical solution to both the reference model as well as the model problem in (19), which we indeed do

possess. However, the purpose of the model is to verify that both the DG simulations as well as the corresponding delay calculations agree with those of the stochastic model. Hence, we introduce notation to explain how the DG calculations can be used to approximate the delay calculations given above. A comparison of the true delay derived using $z(0.5, t)$ is then compared with the delay computed using the DG approximations.

The DG calculation for z_h is used to approximate $s(t)$ as defined in (25). The notation s_h represents the approximation to $s(t)$ calculated using $z_h(0.5, y)$ in equation (25) and applying a numerical quadrature rule to approximate the integral. The $N(t)$ calculation is done using the true solution of equation (19) with $\beta(x, t) = 1$ for all x and t and the initial and boundary conditions given by z_0 prescribed in equations (20) and (21).

Using the DG calculations, the approximation of the average delay is defined as

$$D_G(T_0, T_1) = \frac{1}{\int_{T_0}^{T_1} f(z_h(0.5, t)) dt} \int_{T_0}^{T_1} f(z_h(0.5, t))(s_h(t) - t) dt, \quad (27)$$

where a composite Trapezoidal rule is used to approximate the integrals. The composite Trapezoidal rule is chosen because it uses a low order approximation to the integrand. This is not necessary in the model problems that we investigate. However, in the more biologically meaningful models, one finds that the integrand is highly oscillatory and non-differentiable, and in such situations, less error is introduced by using a low-order composite scheme.

5.3.3 Comparison of the Delay Calculations

For the example included here, we assume that the DNA strand is 1000 nucleotides in length, and we compute the pause-induced delay per affected polymerase induced by a single pause. The pause is positioned in the middle of the DNA strand ($x = 0$) and is active in the time interval $[1000, 1020]$ seconds. We compute the delay over the time interval $[T_0, T_1] = [950, 1250]$ at a position $x = 0.1$, which corresponds to 100 basepairs downstream from the pause, and at the end of the DNA strand ($x = 0.5$). For the delay computed at $x = 0.1$ we have

$$D_G(950, 1250) = 1.996 \text{ sec/RNAP}, \quad D_S(950, 1250) = 1.716 \text{ sec/RNAP}, \quad (28)$$

and for the delay computed at $x = 0.5$ we have

$$D_G(950, 1250) = 4.318 \text{ sec/RNAP}, \quad D_S(950, 1250) = 3.517 \text{ sec/RNAP}. \quad (29)$$

First observe that the delay is larger further away from the pause (at $x = 0.5$), since the aggregation of dense traffic continues to affect the new polymerases that encounter it, and it continues to increase the delay of polymerases within the group. We note that in spite of the difference in flux between the PDE

and stochastic simulation along the shock in the PDE (Figure 2), the delay computed from the stochastic model is 85% and 81% of the PDE computed delay, at $x = 0.1$ and $x = 0.5$ respectively. This encourages us to use the more efficient PDE solution to estimate the induced delay for a biologically realistic number and distribution of pauses in the next section.

6 Delay Computations for a Ribosomal RNA Model

In this section, we focus on the quantitative aspects of the PDE model that can be used to test hypotheses about a problem in cell biology. We examine ribosomal RNA transcription for the *rrn* operon which is 5450 nt long and has been studied experimentally in Bremer and Dennis 1996. According to experimental data, this gene requires about 60 seconds to transcribe. If one assumes that the elongation rate of each polymerase is constant and the polymerase experiences no crowding or pauses, then this results in the average transcription speed of 91 nt/s. This speed is about twice the mRNA transcription speed (Condon et al. 1993, Vogel and Jensen 1995, Vogel and Jensen 1997, Dennis et al. 2009). This difference stems from the *rrn* operon-specific modifications of RNAP, which require both the presence of anti-termination sequences near the *rrn* operon leader region and the presence of interacting proteins. The data suggest that the RNAP modifications allow read-through of Rho-dependent terminators (Albrechtsen et al. 1990, Vogel and Jensen 1995, Vogel and Jensen 1997, Zellars and Squires 1999), but these modifications may also decrease polymerase ubiquitous shorter pauses (Vogel and Jensen 1995, Yang and Roberts 1989, Burns et al. 1998). The mechanism by which these modifications can influence duration and/or frequency of ubiquitous pauses is not completely understood (Neuman et al. 2003, Landick 2009, Galburt et al. 2007). Using our model we test the hypothesis that the anti-termination complex affects duration and/or frequency of ubiquitous pauses.

Taking into consideration observed density of polymerases on *rrn* we assume that the polymerase reads through all Rho-dependent terminators. However, we assume that it encounters the density of ubiquitous pauses that has been observed at other, non *rrn* sequences. We then simulate the PDE model of transcription for a biologically relevant range of raw elongation velocities β (ranging from $\beta = 90$ to $\beta = 220$ nt/second) and observe whether the measured average crossing time matches the observed crossing time of 60 seconds.

According to Condon et al. (1993), there are on average 53.4 RNAP on the *rrn* operon. With a length of 32 nt for each RNAP (Krummel and Chamberlin 1989); this gives an estimate of the fraction of the DNA strand that is covered to be approximately $(53.4 \times 32)/5450 = 0.31$. That is, on average, approximately 31% of the DNA strand is occupied at any given time. We note that the elongation speed on an empty *rrn* operon has not been measured; in the previous paragraph, the estimated speed of 91 nt/s is based on an experimental observation at a full density of polymerases and in the presence of pauses. Using the assumptions of the PDE model, one can compute the raw elongation

velocity β , under the condition that no pauses are present and assuming that the measured average elongation speed of 91 nt/s is a result of polymerase crowding. Indeed the steady state velocity (at density $z_0 = 0.31$) is equal to raw elongation velocity multiplied by $1 - z_0$. This gives an estimate for the raw elongation velocity as

$$\beta = 91/(1 - 0.31) = 132 \text{ nt/s},$$

and this indicates that accounting for the phenomena of polymerase crowding would decrease a raw elongation speed of 132 nt/s for a single RNAP to the observed speed of 91 nt/s.

We now use the model to compute the effect of pauses on the observed transcription speed. As far as we know, the number and frequency of pauses were not measured for rRNA transcription, but they were measured for transcription of other genes. There the frequency of pauses is $0.1s^{-1}$ (Dennis et al. 2009) which means there are about 6 pauses encountered per minute of elongation. To test the hypothesis that the association of an anti-termination complex with the RNAP transcribing *rrn* operon is not only causing read-through of the Rho-dependent terminators but is also suppressing ubiquitous pauses, we assume that the RNAP on *rrn* encounters 6 pauses per minute.

In order to account for polymerase crowding and the effect of pauses, we consider the PDE model (19)-(21) and the incorporation of a large number of pauses at randomly chosen nucleotide locations along the DNA strand. The location and duration of each pause is then encoded into the PDE by constructing $\beta(x, t)$ to account for these pauses in the same way that the one pause is incorporated into the model problem in equation (22). DG is used for the numerical simulation of the PDE model, and the discretization is constructed so that each mesh element represents one nucleotide location. We simulate *rrn* transcription for the gene mentioned above, which is 5450 nt long. The spatial domain $[-0.5, 0.5]$ is partitioned into 5450 finite elements of uniform length, and we consider the spatial-temporal product space when determining the placement of the pauses. The spatial locations for the pauses are chosen uniformly from the elements numbered from 1 to 5450, with the exception that the algorithm is modified so that a pause location is not chosen to be in the first or last element of the mesh. The pause durations are chosen so that, on average, forty percent of the pauses are long pauses and sixty percent of them are short pauses. The duration of time for each of these types of pauses is chosen according to an exponential distribution. Using experimental data from Neuman et al. 2003, the short pauses are exponentially distributed with a mean of approximately 1.2 seconds, and the long pauses are exponentially distributed with a mean of approximately 6 seconds. The mean length of a pause is $\tau = 3.12$ seconds.

The numerical simulations are constructed to be consistent with Dennis et al. (2009) and Klumpp and Hwa (2008) where, on average, a polymerase encounters 6 pauses per minute of elongation. In order to determine the total number of pauses required in a given computational domain, we consider a polymerase starting at an arbitrary time $t \in [0, T]$ transcribing a DNA of

fixed length (with T measured in seconds). For this calculation we can assume without loss that the elongation speed of the polymerase on a free DNA is 1. Select a single pause of length τ uniformly from a region $S := \{(x, t) \mid x \leq t \leq x + T\}$, which represents all positions and times that are reachable by a polymerase initiating at the start of the DNA strand at a time $t \in [0, T]$. The probability that a single polymerase starting at arbitrary time $t \in [0, T]$ will experience the pause is τ/T . Therefore, if we select $\frac{T}{\tau}$ pauses of average length τ uniformly from S , we expect that any given polymerase will experience a pause. Since hitting two pauses consecutively are independent events, if we want to simulate the situation where, on average, a polymerase encounters 6 pauses per minute, we must select a total number of $6\frac{T}{\tau}$ pauses.

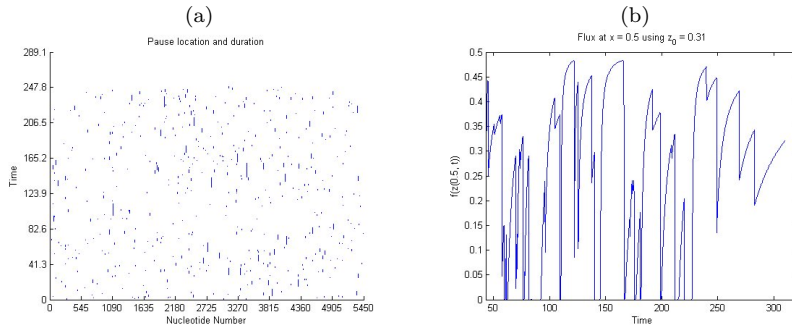


Fig. 3 (a) Plot illustrating the pause location and durations (given in terms of the nucleotide numbers and dimensional time in seconds) used in the PDE simulation. (b) The DG approximation of flux at the termination site $x = 0.5$ as a function of time, using the initial density of $z_0 = 0.31$. Note the oscillatory behavior in the flux function that is caused by the randomness of pause locations and durations. Flux functions of this type are integrated numerically in order to compute the pause-induced delay per polymerase.

With $\tau = 3.12$ seconds, and our selection of the time scaling so that computational domain where the delay is measured corresponds to a simulation of approximately 5 minutes of the transcription process, a sample of the results can be seen in Figure 3. Part (a) of Figure 3 gives a graphical representation of the location and time duration of the pauses incorporated into the simulation. Figure 3(b) is an illustrative sample of a DG simulation of the flux at the termination site over the relevant time interval, that is, it gives a sample of the type of function $f(z_h(0.5, t))$ used for the delay calculation as described in equation (27). For the calculations presented in this paper, we choose T_0 and T_1 for the delay calculation so that it represents an interval of time where the integrals in (27) measure the output of the system once it has already experienced a randomly-distributed number of pauses.

Employing a large number of simulations of the type described in the previous paragraphs, we test the hypothesis that increasing the raw elongation speed in the presence of pauses can lower the average crossing time. The results are shown in Figure 4 and Table 1. Note that a coarser grid of 2725

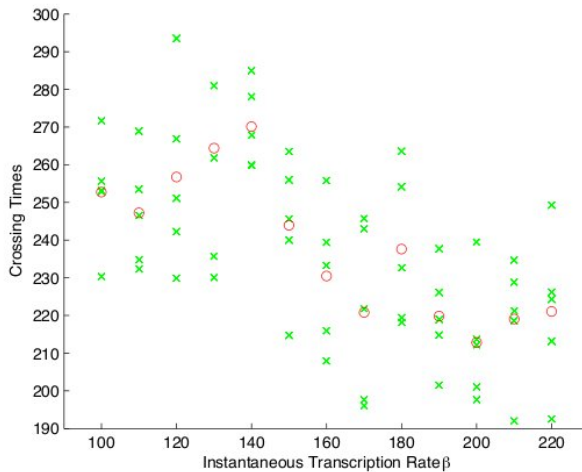


Fig. 4 Computation of an average crossing time (vertical axis) of a polymerase on the *rrn* operon. For each value of raw elongation speed β we simulate 5 minutes of transcription for 5 random pause distributions (marked by crosses), which are consistent with the observed pause density (see the text). The average value of the five trials is indicated by a circle.

elements was used for the data presented here. That is, the discretization is constructed so that each mesh element represents two nucleotide locations. We see that even at $\beta = 220$ the average crossing time per polymerase is almost 4 minutes. The data in Figure 4 suggests that when the polymerase experiences a number of ubiquitous short pauses, increasing elongation speed is insufficient to lower total crossing times for each polymerase by a significant amount. This suggests that in order to approach the experimentally observed crossing time of approximately 60 seconds, the polymerase modifications must suppress sequence-independent ubiquitous pauses. Furthermore, this supports the hypothesis that the anti-termination complex can significantly increase the transcription rate of rRNA, which is a key growth limiting factor in *E. coli*.

We also note that the crossing time does not monotonically decrease with increasing β , and it seems to level off at high β , perhaps approaching an asymptote that is much higher than 60 seconds. The asymptotic behavior is perhaps expected, as the crossing time should be limited by the presence of a critical number of pauses. However, the non-monotonicity of the crossing time suggests that multiple factors may influence the resulting crossing time. In particular, the pauses are located at a wide variety of spatial locations within the domain, and we believe that interaction between RNAPs that encounter multiple pauses at closely related spatial locations can affect the overall crossing time in complex ways. As a first step at understanding this phenomena, the following section includes a preliminary study of how the crossing time varies with respect to parameters dictating pause location and pause duration.

Table 1 Mean delay and mean overall crossing time per polymerase in the presence of ubiquitous pauses and various values of β . DG simulation with a spatial grid of 2725 elements was used.

Raw Elongation Velocity β	Mean Delay per polymerase (secs)	Mean Total Elong. Time (secs)
100 nt/s	198 s	253 s
110 nt/s	197 s	247 s
120 nt/s	211 s	257 s
130 nt/s	222 s	264 s
140 nt/s	231 s	270 s
150 nt/s	208 s	244 s
160 nt/s	196 s	230 s
170 nt/s	189 s	221 s
180 nt/s	207 s	238 s
190 nt/s	191 s	220 s
200 nt/s	186 s	213 s
210 nt/s	193 s	219 s
220 nt/s	196 s	221 s

7 Parameter Studies for Two Simple Models

To understand the source of non-monotonicity mentioned above, we summarize results from two simple model problems.

7.1 One Pause Model

Returning to the one-pause model discussed in Section 5.1.2, we show that the crossing times are predictably monotone as a function of pause location. Using equations (19)-(21) with β parameterized as in (22) for the case where $\zeta = 0.2$ and $\xi = 0.3$, the spatial location of the single pause is varied across most of the interior of the domain, see Figure 5. We see that the pause located close to the beginning of the strand increases the average crossing time more than a pause of the same duration positioned near the end of the strand. The actual elongation speed here is somewhat irrelevant; however, we have used an elongation speed that results in a crossing time that is biologically relevant to the model analyzed in this paper. We have also confirmed the intuition that longer pauses increase elongation time more than short pauses; we omit these results for brevity.

7.2 Two Pause Model

The introduction of a second pause with a parametrized location is enough to illustrate the complex behavior of the system in the presence of multiple pauses. With two pauses of the same time duration positioned relatively close

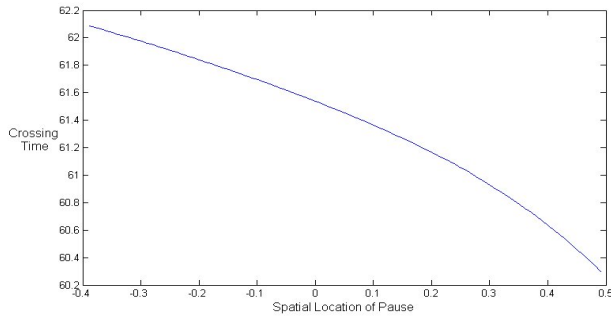


Fig. 5 Plot representing the monotone decrease in average crossing time as the spatial location of one pause is varied from $x = -0.4$ to $x = 0.49$. Simulations were carried out using the DG framework (described in previous sections) with a background density of $z_0 = 0.31$ with 1000 mesh elements.

together, numerical simulations indicate that the average crossing time exhibits complex behavior that includes both local maxima and minima. The first pause (the one occurring earlier in time) is fixed in the center of the domain at $x = 0$ (representing the middle of the DNA strand), begins at a time 6 seconds into the simulation and has a duration of approximately 3.2 seconds. The second pause begins at a time of about 35.9 seconds. The spatial location of this second pause is varied across the domain, and we plot the average crossing time as a function of this spatial location.

We observe a non-monotone behavior in the crossing time for this situation, as seen in Figure 6(a). Examining the contour plots in (b-d), it is visible that the first local minimum crossing time happens for the parameter location where the shock wave from the first pause tangentially intersects the spatial position of the second pause at the time when the second pause begins, see Figure 6 (b). This allows the rarefaction wave resulting from the release of the first pause to spread into the area of low density created by the second pause; thereby allowing the RNAPs that are not stopped by the second pause to speed up as they continue to elongate across the DNA strand. In addition, there is a local maximum observed in Figure 6(a) which corresponds to the situation shown in the contour plot in Figure 6(c), where many of the RNAPs affected by the first pause are also affected by the traffic jam created by the second pause. This corresponds to the intersection of the characteristic created by the release of the second pause with the shock emanating from the first pause, separating polymerases affected by the first pause from those unaffected by it. Finally, in Figure 6(d) we show the situation at the second local minimum. Here the second pause is on the border of the shadow of the first pause and hence very few polymerases are affected by it since the density of traffic in that region is relatively low (because of the first pause).

The simulations of these simple models suggest that the complex interaction between polymerase density and multiple pauses may lead to non-monotonicity in crossing times for any given choice of raw elongation speed β .

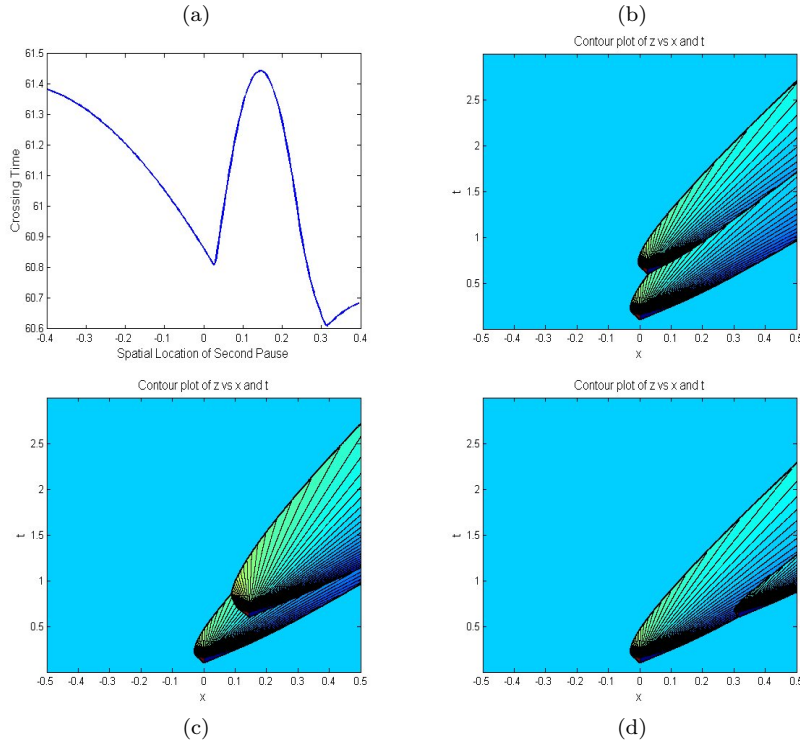


Fig. 6 (a) Expresses average crossing time as a function of the spatial location of the second pause (b) Contour plot of the PDE model simulation of density $z(x, t)$ where the location of the second pause corresponds to the first local minimum value of the crossing time. (c) Density and location of the second pause at local maximum of the crossing time, (d) density and location at the second local minimum.

While we cannot rule out that the non-monotonicity in Figure 4 results from an insufficient number of samples, Figure 6 suggests that it is plausible that it may arise from the interplay between number and position of pauses, density of RNAPs and the speed β .

8 Conclusions

In this paper we rigorously derived a nonlinear conservation law that describes the motion of a molecular polymerization machine on a one-dimensional ribbon. Our model has roots in a stochastic TASEP process, which gives rise to an ODE model describing mean occupancy at each position along the ribbon. Using the theory of convergence of finite volume methods for nonlinear conservation laws, we show that a time discretization of this ODE model can be used to build an approximate solution of a nonlinear conservation law. Increasing the number of ODEs and refining the time discretization yields a sequence of approximations that weakly converge to a solution of the conservation law.

We apply our PDE model to the gene transcription process and numerically investigate the effect of short ubiquitous pauses on transcription efficiency. We compute average crossing time as well as pause-induced delay per elongating polymerase and compare these estimates between the PDE model and the stochastic TASEP model.

Finally, we address a biologically relevant question related to the transcription of the *rrn* gene in *E.coli*. A transcript of this gene forms a core of each ribosome. At times of rapid growth, demand for production of ribosomes requires high transcription rate of the *rrn* gene. Multiple polymerases transcribe the gene at the same time, and this high density in the presence of pauses makes polymerase traffic jams likely. It is known that formation of a so called anti-termination complex doubles the average transcription speed by primarily allowing read-through of several terminator regions within the *rrn* gene. It is not clear how much the formation of the anti-termination complex affects duration and number of ubiquitous short pauses during the transcription. Using our model we show that if we assume that the elongating polymerase encounters the same number of short pauses as were measured on regular genes, then the crossing time remains well above the experimentally observed 60 sec for a range of realistic raw elongation velocities. Our results suggest that the anti-termination complex must act to shorten the duration or to decrease the number (or both) of the short ubiquitous pauses.

Acknowledgment T. G. was partially supported by NSF grant DMS-0818785, NSF CMMI grant 0849433 and NIH R01 grant 1R01AG040020-01; J. G. was partially supported by an undergraduate summer research grant from Montana State University. L.D. and J.T. were partially supported by NSF grant DMS-1226213. The authors wish to thank the reviewers of this manuscript for their careful reading and constructive suggestions. Their comments helped the authors to significantly improve the finished product.

References

1. Albrechtsen, B., Squires, C., Li, S., Squires, C.: Antitermination of characterized transcription terminators by the *escherichia coli* *rrn* leader region. *J. Mol. Biol.* **213**, 123–134 (1990)
2. Andjel, E.D., Kipnis, C.: Derivation of the hydrodynamical equation for the zero-range interaction process. *Annals of Probability* **12**(2), 325–334 (1984)
3. Andjel, E.D., Vares, M.E.: Hydrodynamic equations for attractive particle systems on \mathbb{Z} . *J Stat Phys* **47**(1/2) (1986)
4. Argall, B., Cheleshkin, E., Greenberg, J.M., Hinde, C., Lin, P.J.: A rigorous treatment of a follow-the-leader traffic model with traffic lights present. *SIAM J. Appl. Math.* **63**(1), 149–

- 168 (electronic) (2002). DOI 10.1137/S0036139901391215. URL <http://dx.doi.org/10.1137/S0036139901391215>
5. Arnold, D.N., Brezzi, F., Cockburn, B., Marini, L.D.: Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.* **39**(5), 1749–1779 (2001/02). DOI 10.1137/S0036142901384162. URL <http://dx.doi.org/10.1137/S0036142901384162>
 6. Aw, A., Klar, A., M., Materne, T.: Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM Journal on Applied Mathematics* **63**(1), 259–278 (2002). DOI 10.1137/S0036139900380955. URL <http://link.aip.org/link/?SMM/63/259/1>
 7. Aw, A., Rascle, M.: Resurrection of “second order” models of traffic flow. *SIAM J. Appl. Math.* **60**(3), 916–938 (electronic) (2000). DOI 10.1137/S0036139997332099. URL <http://dx.doi.org/10.1137/S0036139997332099>
 8. Bai, L., Shundrovsky, A., Wang, M.D.: Sequence-dependent kinetic model for transcription elongation by rna polymerase. *Journal of Molecular Biology* **344**(2), 335–49 (2004)
 9. Barkema, G., Marko, J., Widom, B.: Electrophoresis of charged polymers: Simulation and scaling in the repton model. *Phys. Rev. E* **49**, 5303–09 (1994)
 10. Bellamo, N., Dogbe, C.: On the modelling of traffic and crowds: a survey of models, speculations, and perspectives. *SIAM Review* **53**(3), 409–463 (2011)
 11. Bremer, H., Dennis, P.: *Escherichia coli* and *Salmonella typhi*-*murium*: Cellular and Molecular Biology, chap. Modulation of chemical composition and other parameters of the cell by growth rate, p. 421. ed Neidhardt EA (Sinauer, Sunderland, MA) 2nd Ed., College Station, Texas (1996)
 12. Burns, C., Richardson, L., Richardson, J.: Combinatorial effects of *nusA* and *nusG* on transcription termination and rho-dependent termination in *escherichia coli*. *J Mol Biol* **278**, 307–316 (1999)
 13. Chou, T., Mallick, K., Zia, R.K.P.: Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport. *Rep. Progr. Phys.* **74**(11), 116,601, 41 (2011). DOI 10.1088/0034-4885/74/11/116601. URL <http://dx.doi.org/10.1088/0034-4885/74/11/116601>
 14. Cockburn, B., Shu, C.W.: Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *J. Sci. Comput.* **16**(3), 173–261 (2001). DOI 10.1023/A:1012873910884. URL <http://dx.doi.org/10.1023/A:1012873910884>
 15. Condon, C., French, S., Squires, C., Squires, C.L.: Depletion of functional ribosomal rna operons in *escherichia coli* causes increased expression of the remaining intact copies. *The EMBO Journal* **12**(11), 4305–4315 (1993)
 16. Daganzo, C.F.: Requiem for second-order fluid approximations of traffic flow. *Transportation Res. Part B* **29B**(4), 277–286 (1995)
 17. Dennis, P.P., Ehrenberg, M., Fange, D., Bremer, H.: Varying rate of rna chain elongation during *rrn* transcription in *escherichia coli*. *Journal of*

- Bacteriology **191**(11), 3740–3746 (2009)
18. Derrida, B., Evans, M., Hakim, V., Pasquier, V.: Exact solution of a 1d asymmetric exclusion model using a matrix formalism. *J. Phys. A: Math. Gen.* **26**, 14931517 (1993)
 19. Derrida, B., Lebowitz, J.L., Speer, E.R.: Shock profiles for the asymmetric simple exclusion process in one dimension. *J Stat Physics* **89**(1/2), 1–33 (1997)
 20. Ferrari, P.A., Kipnis, C., Saada, E.: Microscopic structure of travelling waves in the asymmetric simple exclusion process. *Annals of Probability* **19**(1), 226–244 (1991)
 21. Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., Bustamante, C.: Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner. *Nature* **446**(7137), 820–3 (2007)
 22. Gazis, D.C., Herman, R., Rothery, R.W.: Nonlinear follow-the-leader models of traffic flow. *Operations Res.* **9**, 545–567 (1961)
 23. Haberman, R.: Mathematical models, *Classics in Applied Mathematics*, vol. 21. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1998). Mechanical vibrations, population dynamics, and traffic flow, An introduction to applied mathematics, Reprint of the 1977 original
 24. Harris, R.J., Stinchcombe, R.B.: Disordered asymmetric simple exclusion process: Mean-field treatment. *Phys. Rev. E* **70**, 016,108 (2004). DOI 10.1103/PhysRevE.70.016108. URL <http://link.aps.org/doi/10.1103/PhysRevE.70.016108>
 25. Heinrich, R., Rapoport, T.: Mathematical modelling of translation of mrna in eucaryotes: Steady states, time-dependent processes and application to reticulocytes. *J. Theor. Biol.* **86**, 279–313 (1980)
 26. Hesthaven, J.S., Warburton, T.: Nodal discontinuous Galerkin methods, *Texts in Applied Mathematics*, vol. 54. Springer, New York (2008). DOI 10.1007/978-0-387-72067-8. URL <http://dx.doi.org/10.1007/978-0-387-72067-8>. Algorithms, analysis, and applications
 27. von Hippel, P.H.: An integrated model of the transcription complex in elongation, termination, and editing. *Science* **281**(5377), 660–5 (1998)
 28. Janowsky, S.A., Lebowitz, J.L.: Finite-size effects and shock fluctuations in the asymmetric simple-exclusion process. *Phys. Rev. A* **45**, 618–625 (1992). DOI 10.1103/PhysRevA.45.618. URL <http://link.aps.org/doi/10.1103/PhysRevA.45.618>
 29. Kandel, D., Mukamel, D.: Defects interface profile and phase transitions in growth models. *Europhys. Lett.* **20**, 325–9 (1992)
 30. Kireeva, M.L., Kashlev, M.: Mechanism of sequence-specific pausing of bacterial rna polymerase. *PNAS* **106**(22), 8900–5 (2009)
 31. Klumpp, S.: Pausing and backtracking in transcription under dense traffic conditions. *J. Stat. Phys.* **142**, 1251–1267 (2011)
 32. Klumpp, S., Hwa, T.: Stochasticity and traffic jams in the transcription of ribosomal rna: Intriguing role of termination and antitermination. *PNAS* **105**(47), 18,159–164 (2008)

33. Klumpp, S., Lipowsky, R.: Phase transitions in systems with two species of molecular motors. *Europhys. Lett.* **66**, 90–96 (2004)
34. Kolomeisky, A.: Asymmetric simple exclusion model with local inhomogeneity. *J Phys A* **31**, 1153–1164 (1998)
35. Krummel, B., Chamberlin, M.: Rna chain initiation by escherichia coli rna polymerase-structural transitions of the enzyme in early ternary complexes. *Biochemistry* **28**, 7829–7842 (1989)
36. Landick, R.: Transcriptional pausing without backtracking. *PNAS* **106**(22), 8797–8 (2009)
37. LeVeque, R.: *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, New York, NY (2002)
38. Lighthill, M., Whitham, G.: On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A*, **229**(1178), 317–345 (1955)
39. MacDonald, C.T., Gibbs, J.H.: Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers* **7**, 707–725 (1969)
40. MacDonald, C.T., Gibbs, J.H., Pipkin, A.C.: Kinetics of biopolymerization on nucleic acid templates. *Biopolymers* **6**, 1–25 (1968)
41. Mier-y-Teran-R, L., Silber, M., Hatzimanakatis, V.: The origins of time-delay in template bio-polymerization processes. *PloS Comput Biology* **6**(4), e1000726 (2010)
42. Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., Block, S.M.: Ubiquitous transcriptional pausing is independent of rna polymerase backtracking. *Cell* **115**(4), 437–447 (2003)
43. Newell, G.F.: Nonlinear effects in the dynamics of car flowing. *Operations Res.* **9**, 209–229 (1961)
44. Rascle, M.: An improved macroscopic model of traffic flow: derivation and links with the Lighthill-Whitham model. *Math. Comput. Modelling* **35**(5-6), 581–590 (2002). DOI 10.1016/S0895-7177(02)80022-X. URL [http://dx.doi.org/10.1016/S0895-7177\(02\)80022-X](http://dx.doi.org/10.1016/S0895-7177(02)80022-X). Traffic flow—modelling and simulation
45. Schutz, G., Domany, E.: Phase transitions in an exactly soluble one-dimensional exclusion process. *J. Stat. Phys.* **72**, 277–96 (1993)
46. Shaw, L., Zia, R., Lee, K.: Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Phys. Rev. E* **68**, 021,910 (2003)
47. Spitzer, F.: Interaction of Markov processes. *Advances in Math.* **5**, 246–290 (1970) (1970)
48. Tadigotla, V., Maoileidigh, D.O., A.Sengupta, Epshtein, V., Ebright, R., Nudker, E., Ruckenstein, A.: Thermodynamic and kinetic modeling of transcriptional pausing. *PNAS* **103**(12), 4439–4444 (2006)
49. Tripathy, G., Barma, M.: Driven lattice gases with quenched disorder: Exact results and different macroscopic regimes. *Phys. Rev. E* **58**, 1911–1926 (1998). DOI 10.1103/PhysRevE.58.1911. URL <http://link.aps.org/doi/10.1103/PhysRevE.58.1911>

50. Vogel, U., Jensen, K.: Effects of the antiterminator box a on transcription elongation kinetics and ppgpp inhibition of transcription elongation in escherichia coli. *J. Biol. Chem.* (270), 18,335–18,340 (1995)
51. Vogel, U., Jensen, K.: Nua is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated rna and mrna. *J. Biol. Chem.* (272), 12,265–12,271 (1997)
52. Wick, D.: A dynamical phase transition in an infinite particle system. *J Stat Physics* **38**(5/6), 1015 (1985)
53. Yang, X., Roberts, J.: Gene q antiterminator proteins of escherichia coli phages 82 and lambda supress pausing by rna polymerase at rho-dependent terminators and other sites. *Proc Natl Acad Sci USA* **86**, 5301–5305 (1989)
54. Zellars, M., Squires, C.: Antiterminator-dependent modulation of transcription elongation rates by nusB and nusG. *Mol. Microbiol.* (32), 12961304 (1999)
55. Zhang, P., Liu, R.X.: Generalization of Runge-Kutta discontinuous Galerkin method to LWR traffic flow model with inhomogeneous road conditions. *Numer. Methods Partial Differential Equations* **21**(1), 80–88 (2005). DOI 10.1002/num.20023. URL <http://dx.doi.org/10.1002/num.20023>
56. Zia, R.K.P., Dong, J.J., Schmittmann, B.: Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *J. Stat. Phys.* **144**(2), 405–428 (2011). DOI 10.1007/s10955-011-0183-1. URL <http://dx.doi.org/10.1007/s10955-011-0183-1>