21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Parameter Studies for a Nonlinear Continuum Model of Transcription*

L. Davis[1], T. Gedeon[2] and J. Thorenson[3]

Abstract— **A discontinuous Galerkin finite element method is used to simulate solutions to a nonlinear PDE used to model the biological process of transcription; the process by which RNA polymerase transcribe the genetic information from DNA and transfer it to mRNA. The transcription process is punctuated by short, frequent RNAP pauses, and these pauses cause a delay in the total time duration of the transcription process. The DG solution to the nonlinear model is used to calculate the delay and to quantify the effects of the pauses on the overall transcription time. Preliminary parameter studies indicate that in a system with multiple pauses both the location and time duration of the pauses can significantly effect the average delay experienced by an RNAP.**

I. INTRODUCTION

Protein synthesis is a complex biological process that requires an immense amount of cellular energy as well as coordination and regulation of hundreds of molecules. The two stages of protein synthesis are *transcription* and *translation*. These two processes form the crucial components in the transfer of genetic information from DNA to protein. Transcription involves the transfer of the genetic information from the DNA to messenger mRNA. In bacteria, RNA polymerases (RNAP) translocate along the DNA strand and transcribe the DNA's genetic code into the mRNA molecule. Translation is the process by which ribosomes translocate along a mRNA strand and produce proteins. The ribosomes provide the so-called assembly machinery which uses the genetic information contained in the mRNA to construct a polypeptide chain which subsequently folds to form a protein. Here we consider a simple model that includes only a few of the fundamental underlying mechanisms of the transcription process. The simplified mathematical model is considered as an attempt to gain insight into the limiting behavior of the protein production process under certain cellular conditions.

Both transcription and translation consist of the motion of a complex machine along a one-dimensional strand. The motion has roughly four parts: assembly of the machinery at the start of the strand, movement initiation, elongation and termination. In this paper, we concentrate on the process of transcriptional elongation. It has been observed experimentally that the motion of RNAP is not uniform. Single molecule observation using optical traps indicate that the elongating polymerase transcribes rapidly but frequently pauses, and the duration of the pauses is roughly bi-modal [3] with means 1.2 ± 0.1 sec with amplitude 60% and 6 ± 0.4 sec with amplitude 40%.

The existence of transcriptional pauses brings up the possibility that at high transcription initiation rates, a paused polymerase can prevent the forward motion of one or more polymerases that follow it, thus creating traffic jams [6]. While the density of polymerases on most genes is probably not sufficiently large to produce traffic jams, there are special genes, where the density of polymerases is very high. As an example, transcription of the ribosomal rRNA accounts for over half of the total transcriptional activity in Escherichia coli in high growth conditions [4], 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 [5]. To sustain the high cellular demand on ribosome RNA synthesis, the density of polymerases on the rrn operon is very high. In [1], three types of models are considered in order to quantify the combined effect of high polymerase density and ubiquitous pauses on the transcription rate of ribosomal RNA. One of the three models is a continuum model taking the form of a hyperbolic conservation law with a nonlinear flux component. This paper focuses on one aspect of the numerical simulations and results related to that continuum model. An advantage of using this particular nonlinear PDE is that RNAP pauses at bottleneck nucleotide locations along the DNA strand can be modeled in a very simple manner, and closed form solutions can be written for several simple model problems of interest. This allows us to validate our simulation results prior to tackling the more relevant biologically motivated model formulations, see [14], [2] for many details related to this validation.

As noted in both [1] and [2], numerical simulations of the model for biologically relevant distributions of transcriptional pauses indicate a large amount of variability in the resulting average transcriptional time per polymerase. With this in mind, parameter studies are presented to show that both the location and the time duration can significantly affect the amount of delay experienced by RNAP transcribing along the DNA strand. We present a study of the behavior of the calculated transcriptional delay per polymerase for various configurations of the pauses within the nonlinear model, and these results are used as a means to analyzing the fundamental biological questions related to the transcription process. A larger goal of the research is to gain insight into the effects these pauses have on total throughput of polymerases within the system, and the results of this study allow us to identify various parameterizations of the pauses that can illustrate the complex nature of such a system. Model simulations are carried out using a Discontinuous Galerkin formulation, and the following section begins with a brief discussion of the particular implementation of the DG method used here, including its advantages and its limitations. The fundamental framework of the DG approach for a nonlinear PDE in con-

ISBN: 978-90-367-6321-9

servation law form is considered. Solutions of the PDE model exhibit discontinuities as well as corners that advect as the time variable increases. These discontinuities correspond to the location of transcriptional pauses along the DNA strand, and the corners appear as the rarefaction waves that sweep through the domain once the transcriptional pause has ended. Hence, the numerical scheme is constructed with the intent of accurately capturing this behavior. In [2] several formulations of the model problem are considered where errors in the numerical approximations are measured, and we observe the correct order of convergence away from the spatial location of the corners in the solution. Moreover, the corners in the solutions are accurately tracked by the numerical method as they propagate through the spatial domain over time, and spurious oscillations are minimized by the use of appropriate numerical techniques referred to as slope limiters. In this paper, we focus on a brief description of the mathematical model, the incorporation of the transcriptional pauses into the model and the parameter studies that allow us to explore the fundamental biological questions mentioned above.

II. DISCONTINUOUS GALERKIN SCHEME FOR MODEL SIMULATIONS

The Discontinuous Galerkin finite element method (DG) is a numerical method designed to combine the tools of numerical flux functions and slope limiters from the finite volume method [10] with the geometric flexibility and high-order accuracy of the finite element method [7], [8]. Originally developed to solve PDEs describing conservation laws, the method is characterized by a discretization of the spatial domain of the PDE into elements and a local approximation of the weak solution to the PDE on each element resulting in a local system of ODEs. This local system is then solved by a high-order total variation diminishing (TVD) Runge-Kutta method. Finally a numerical flux function is used to connect the local solutions on the boundaries of the elements. In contrast to traditional finite elements, the weak form is solved locally on each element. The local, and therefore possibly discontinuous, nature of the approximation allows one to capture potential discontinuities at interfaces very accurately, and the use of slope limiters prevents spurious oscillations that can occur when using classical methods.

The choice of numerical flux is one of the key features of DG as its purpose is to connect the local finite element solutions for construction of the global solution and weakly impose any boundary or interface conditions from the PDE. When chosen properly, it also provides stability for the scheme. During the derivation of the weak form, the numerical flux function is introduced after the standard first integration by parts step as an approximation to the flux on the boundary of the elements. For nonlinear scalar conservation laws, the DG formulation is significantly more complicated than in the linear case. Once a nonlinear flux function is introduced to the conservation law, the solution may have discontinuities due to shocks, even though initial and boundary data is continuous. In addition to shocks, artificial oscillations may appear in the DG solution due to how the nonlinear flux function is approximated within the formulation. A slope limiter is used to prevent these oscillations, and it is essential for the stability of the computations. The following section discusses the specific implementation used for the numerical results contained in this paper.

A. Discontinuous Galerkin Scheme for Model Simulations

This section describes the DG method that is used to numerically approximate the solution of a nonlinear PDE in conservation law form

$$z_t + [f(t,x,z)]_x = 0, \quad x \in [L,R], \quad t \geq 0 \tag{1}$$

where $f(t,x,z)$ represents a given flux function, and particular choices for the flux function and specific boundary and initial conditions are discussed in a later section. Here we give a very brief overview of the DG approach, and the interested reader is referred to a wealth of literature on the topic for more details, see [8], [11], [12] and the references contained therein.

Let the function z_h denote the DG approximation of the solution to equation (1) with appropriate boundary and initial data. To obtain this approximation, one first discretizes the spatial domain, $[L,R]$, into $\mathcal{K}$ elements, $D^k = [x_l^k, x_r^k]$ for $k = 1, 2, \cdots, \mathcal{K}$. On each element, choose a set of N interpolation points used to define a basis of Lagrange interpolating polynomials, $\left\{\ell_i^k(x)\right\}_{i=1}^N$. The semi-discrete DG formulation on the k-th element is

$$M^k \frac{d}{dt}\vec{z_h^k} - S^k \vec{f_h^k} = -\left[\vec{\ell^k}(x)(f^*)\right]_{x_l^k}^{x_r^k}$$

where the local mass and stiffness matrices are

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

and f^* is a numerical flux defined at the interface used to connect the local solutions. The notation $\vec{z_h^k}$ denotes the vector of unknown coefficients that determine the DG approximation of z on the k-th element of the mesh obtained by using a grid of size determined by the parameter h. Similarly, the notation $\vec{f_h^k}$ denotes the vector of values interpolating the flux function defined in (1) on the k-th element of the mesh obtained by using a grid of size determined by the parameter h. The numerical flux function, f^*, is defined on the boundary of the elements as an approximation to the true flux, $f(z)$, from the conservation law. It depends on the local solutions at the boundary of the element and all neighboring elements. If z^- represents the local solution on the boundary internal to D^k and z^+ is the local solution on the boundary external to D^k, then numerical fluxes have the form

$$f^* = f^*(z^-, z^+).$$

For the model problems presented here, the Local Lax-Friedrichs flux is used in the implementation of the DG formulation, and it is defined as

$$f^*(z^-, z^+) = \{\{f(z)\}\} + \frac{C}{2}[\![z]\!], \tag{2}$$

and

$$C = \max_{\min(z^-,z^+)\le u\le \max(z^-,z^+)} |f'(z)| \tag{3}$$

where $\{\{f\}\} = \dfrac{f^- + f^+}{2}$ is the average across the interface and $[\![z]\!] = \hat{n}^- z^- + \hat{n}^+ z^+$ is the jump along the outward normal. For the one-dimensional problem, $\hat{n} = \pm 1$.

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 [8] and references therein. Matlab code obtained from the website associated with [8] provided the foundational code which was modified and supplemented to generate the PDE simulations presented in later sections.

In order to have high-order accuracy and stability of the time integration, a total variation diminishing Runge-Kutta method is used, see [9]. To maintain stability, the explicit RK method is combined with a slope limiter that is applied after each stage of the RK function evaluations. For the calculations in this paper, we use a relatively simple slope limiter that relies on the `minmod` function, see [10]. One of the main computational issues with using a slope limiter is the smearing of discontinuities and local extremum and therefore a loss of accuracy in a neighborhood of shocks. Since an artificial oscillation is determined by changes in the sign of the slope, all local maxes and mins are incorrectly identified as oscillations. As shown in [2], this issue causes little concern for our simulations; however, it must be taken into account when convergence studies are employed. All calculations in the current work are obtained using a particular third order Runga-Kutta method with slope limiting that is described more thoroughly in [2] and [14]. The time step, Δt, is chosen to satisfy the CFL condition

$$\Delta t \le \frac{\Delta x}{\max_k |f'(z(x^k,t))|}$$

and the max is computed over all interpolation points chosen in the spatial domain. Using this DG formulation for a scalar nonlinear conservation law, the next section outlines its implementation and simulation results for the nonlinear PDE used to mathematically describe the biological process of transcription.

III. NONLINEAR TRANSCRIPTION MODEL

We examine a nonlinear macroscopic PDE model that was originally considered as a continuum model for traffic flow, see [15], [16]. By specifying the flux function in a variety of different configurations, we model transcription behavior along a single DNA strand where various nucleotides experience transcriptional pauses. For brevity's sake, we note that the original formulation of the model and the details of the nondimensionalization process are contained in [14]. The PDE model used for the numerical simulations in this paper is given in terms of the dimensionless variables and is defined on the domain $x \in (-0.5, 0.5)$ and for $t > 0$ as

$$z_t + [\beta(x,t)(1-z)z]_x = 0 \tag{4a}$$

$$z(x,0) = z_0 \tag{4b}$$

$$z(-0.5,t) = z_0. \tag{4c}$$

Hence, we have replaced the arbitrary flux function in (1) with the particular flux function

$$f(z) = \beta(x,t)(1-z)z,$$

and the term $\beta(x,t)(1-z)$ represents the velocity term within the flux function. This choice represents the assumption that traffic flow velocity of the traffic decreases with increases in the density z. Moreover, the coefficient $\beta(x,t)$ represents a maximum velocity (or maximum speed limit), and it is through the manipulation of this term that we incorporate transcriptional pauses into the model. We construct this coefficient $\beta(x,t)$ in a manner that is determined by the nature of the transcriptional pauses and is defined using the following general form. For $i = 1, 2, \dots M$, define a three-tuple (x_i, t_i, d_i) where x_i denotes the (spatial) nucleotide location of the ith pause, t_i denotes the time at which the pause begins and d_i denotes the time duration of the pause.

$$\beta(x,t) = \begin{cases} 0 & x = x_i \text{ and } t_i < t < t_i + d_i, \\ & \text{for } i = 1, 2, \dots M \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

We also assume that $z_0 < 0.5$ so that the initial density allows the flow to be below its maximum possible value until the RNAPs back up behind the paused nucleotide location.

A. Delay Calculations for the Transcription Model

It is known from empirical observation that during the transcription process, a polymerase frequently pauses along the DNA strand causing a transcriptional delay and affecting the instantaneous transcription speed, see [1], [3], [6], [13]. One goal of the current research is to use numerical simulations of various forms of the PDE model described in (4a)-(4c) using (5) in order to quantify the effect of pauses and subsequent delays on the overall transcription rate of the ribosomal RNA. Beginning with the simple case of one pause, we describe the calculation of the *average delay per polymerase*. This idea easily generalizes to the case of multiple pauses, and a sample calculation of this delay value is given for the setting of multiple pauses at the end of the discussion.

In order to compute the effect of pauses on the average delay, we will consider two cases and then combine them into one calculation. First we describe the *reference case* which characterizes the ideal situation in which there is no pause; that is, the coefficient $\beta(x,t) = 1$ for all (x,t). And the initial and boundary conditions are specified with a certain constant *background* density of z_0. Note that the density $z(x,t) = z_0$ is the solution of such an equation for the reference case; therefore, the reference 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, represented in the dimensionless setting by $x = 0.5$, over the time interval

(T_0, t) is

$$N(t) = \int_{T_0}^{t} f_R dy = (t - T_0) f_R \ . \tag{6}$$

For the second case, we consider the situation where transcriptional pauses are present. That is, we calculate the solution of (4a)-(4c) using (5) which incorporates several predetermined pauses at various spatial locations on the DNA at various times during the simulation. For this case, the number of polymerases that have successfully transcribed the DNA by arriving at the termination site in the time measured from T_0 to the time t is given by

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

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 pauses. To achieve this goal, we define the function $s_h(t)$ so that for each $t > T_0$, the function $s_h(t)$ is the instant of time for which

$$\int_{T_0}^{s_h(t)} f(z_h(0.5, y)) dy = N(t). \tag{7}$$

Here, the notation on the right side refers to the number of RNAPs that have arrived at time t for the reference case described by equation (6). The notation s_h represents the fact that the integral on the left side of (7) is calculated using the DG solution $z_h(0.5, y)$ and applying a numerical quadrature rule to approximate the integral. For the right side of (7), the $N(t)$ calculation is done using the equation given in (6) for the reference case with the initial and boundary conditions given by z_0 prescribed in equations (4b) and (4c).

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

$$D_G(T_0, T_1) = \frac{\int_{T_0}^{T_1} f(z_h(0.5, t))(s_h(t) - t) dt}{\int_{T_0}^{T_1} f(z_h(0.5, t)) dt} \tag{8}$$

where T_0 and T_1 represent arbitrary values of the dimensionless time, see [2] for a more detailed discussion. For the calculations contained here, 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 in this paper. However, in the more biologically meaningful models, one finds that the integrands are highly oscillatory and non-differentiable, and in such situations, less error is introduced by using a low-order composite scheme.

IV. Parameter Studies for Models with Multiple Pauses

Computational results for biologically relevant model formulations are discussed in detail in [1], [2], [14], and it is demonstrated that the stochastic nature of the pause locations and durations results in a large amount of variability in the calculations of the *average transcription time per polymerase* for different realizations of the transcriptional pauses. Hence, a statistical analysis is currently underway for the case of the biologically relevant situation (where a large number of pauses is included in the mathematical model). However, as a preliminary quantitative analysis where one can make use of the PDE model without the need for an *a posteriori* statistical analysis, we demonstrate that even for extremely simple cases, the delay function can display complicated behavior. In [1], the authors give numerical results demonstrating that for the very simple case of one pause, the crossing times are predictably monotone as a function of pause location. That is, a pause located close to the beginning of the DNA strand produces a larger average transcription time per RNAP than that of a pause of the same time duration positioned near the termination end of the strand. However, the following discussion illustrates that the placement and duration of even two pauses can have more interesting effects on the delay function. In the following discussion, we consider two types of parameterizations of the spatial locations and time durations of the pauses.

A. Parameterized Time Between Pauses

Here we consider the situation where two pauses occur at the same spatial location but occur during separate periods of time. We construct the velocity coefficient function so that

$$\beta(x,t) = \begin{cases} 0, & x = 0 \ \text{ and } 0.2 < t < 0.3 \\ 0, & x = 0 \ \text{ and } 0.4 + i(0.1) < t < 0.5 + i(0.1) \\ 1, & \text{otherwise} \end{cases}$$

where $i = 1, 2, \cdots, 58$. Note that these pauses have the same duration time, and the adjustment of the parameter i simply determines the time between the end of the first pause and the beginning of the second pause. We measure the average delay per polymerase as a function of both this parameter i as well as the initial and background density z_0 prescribed in equation (4b). The choice of the parameter z_0 is biologically relevant because the number of transcribing RNAPs on the same gene varies significantly over different genes. It is also known that the number of RNAPs can vary significantly on a single gene under different growth conditions. In particular, the *rrn* gene in E.coli can be transcribed by as many as 50 RNAP's simultaneously, while in a low growth condition this number can be close to zero. Fig. 1 shows that for each choice of background density z_0, there is a minimum delay value (Fig. 2), as the parameter i determining the time between the two pauses is varied from 0.2 to 5.9 in small increments. The delay values generally range from nearly 0 to an average delay of about 0.08 in dimensionless time units. Note that this delay represents a particular amount of delay in dimensional time that depends on both the length of the DNA strand and the maximum elongational velocity that is assumed for the given gene. However, one can calculate a percentage of increase in the dimensionless transcription time as compared with the transcription time for the reference

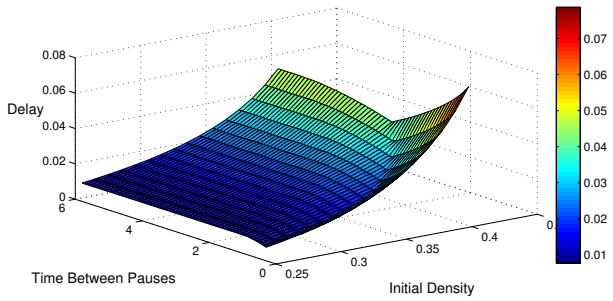

Fig. 1. Surface plot of the average delay for a range of initial density values (4b) and time between the two pauses.

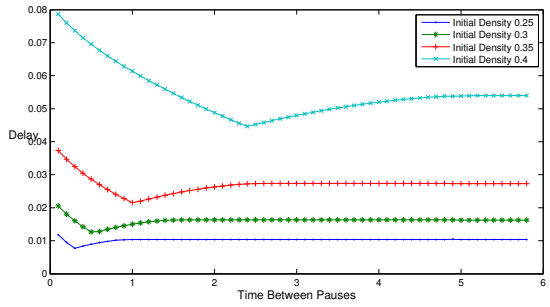

Fig. 2. Cross sections of the average delay plot for several initial density values z_0 in (4b) and time between the two pauses.

case. The transcription time for one RNAP is given by $t^\star = 1/(1 - z_0)$ for the reference case. Then for the range of z_0 chosen here, the transcriptional delay ranges between 0 and 5% of the dimensionless transcription time. Fig. 8 includes the contour plot of the density as a function of both x and t for the simulation corresponding to the minimum delay value using an initial density of $z_0 = 0.3$ from Fig. 2. Upon careful inspection, one observes that the minimum delay occurs when the outer shock caused by the first pause intersects the space and time location of the beginning of the second pause. Intuitively this seems reasonable since the RNAPs that are backed up by the first pause are not affected by the second pause as the backup from the first pause just clears the nucleotide before the second pause begins. Moreover, in the nonlinear model, the elongation velocity of the RNAPs is linearly related to the density, and the velocity of the those affected RNAPs increases with decreasing density. Therefore, the trailing RNAPs that are stopped by the first pause are allowed to elongate at a faster rate during the time interval between the cessation of the first pause and the activation of the second pause.

B. Parameterized Location of a Second Pause

In contrast to the previous section, we now illustrate that if the spatial location of the second pause is varied (but its time duration is fixed), the average delay function can attain both local minimum and local maximum values. For the case shown here, one pause is fixed at $x = 0$ during the times $0.1 < t < 0.15$. The second pause is spatially located at $x = -0.4 + i/200$ for $i = 0, 1, 2, \cdots, 160$ during the times

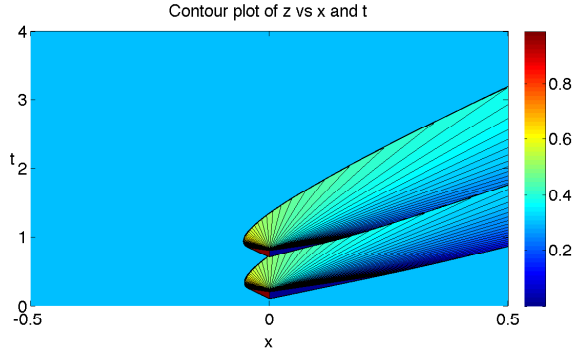

Fig. 3. Contour plot of the PDE simulation corresponding to the minimum delay value in Figure 2 with an initial density value of $z_0 = 0.3$.

$0.6 < t < 0.65$. This includes spatial locations that are behind the fixed pause on the DNA strand as well as some which are located ahead of it on the strand. The average delay was then computed over the time interval (0, 4.1). The right end point of this time interval was chosen so that for each choice of the parameter values, every polymerase affected by the pauses has reached termination and the density is back to the initial value, $z = z_0 = 0.3$ for the current computations. A local maximum and two local minimums in the delay are observed in Figure 4. This indicates that the interaction of the backup caused by these pauses can affect the overall average crossing times of the RNAPs in complex ways, thereby affecting the total number of mRNAs that are produced by the gene in a given time period. Fig. 6 and 7 include the contour plots of density for the simulations corresponding to the parameter values producing local minimum delay values in Fig. 4. These graphs indicate that the minimums occur when the starting time and location of the second pause intersects the shock emanating from the release of the first pause. Figure 5 includes the contour plot of the simulation corresponding to the maximum delay value indicating that it occurs when the shock emanating from the first pause intersects the characteristic created by the release of the second pause. We might also note that the actual percentage of the average delay only ranges from about 0.012 to 0.028 in terms of the dimensionless time variable.

As shown in Figures 2 and 4, both the background density of the RNAPs as well as the pause locations parameterized in time and space have an affect on the average transcriptional delay per polymerase. However, it is unclear which has a larger effect on the average time delay experienced per polymerase. Investigations using these types of parameterizations and a more sophisticated sensitivity analysis are currently underway.

C. Clustering of Pauses

Note that in the above discussion, certain configurations of two pauses are shown to minimize the average delay; while another configuration is shown to result in a maximum delay value. Similar phenomena has been studied in the related mechanism of mRNA translation by Chou and Lakatos [17]. Using a TASEP model with inhomogeneous hopping rates,

the work in that paper demonstrates that the existence of one so-called "defect" or "bottleneck" codon can significantly decrease the elongation speed and thereby decrease protein production. Although the overall framework of the TASEP approach is quite different from the continuum model that we employ in this paper, the modeling of the defect codon in the translational process is similar to our approach of imposing the RNAP pause located at a specific nucleotide in the transcriptional process. Moreover, using the continuum model approach and the limited parameter studies of this section, we are able to draw very similar conclusions to those in [17]. They indicate that a carefully chosen *cluster* of a small number of these defect codons can reduce protein production by at least a factor of two. They also assert that configurations where multiple defect codons are spaced more than a certain number of codons away from each other will not reduce the output of the system any more than that of one single defect codon. The parameter studies contained here for the continuum model lead to similar conclusions; however, the PDE model gives us the ability to parameterize both the spatial location and time occurrences of the pauses. Using this flexibility, we can make observations relating to the short time behavior of the system as opposed to only considering a steady state regime.

As shown in Section IV-A, when two pauses occur at the same nucleotide location and the time between the two pauses is varied, the transcriptional delay values are largest when the pauses occur very close to each other in time. As the time between the two pauses increases, the delay values decrease until a minimum delay value is attained. However, these simulations also indicate that as the time between pauses increases the delay value will exhibit a modest increase before leveling off to a constant value indicating that the two pauses can have only a fixed amount of influence on the average transcriptional delay once they occur further apart in time. The results in Section IV-B are far more interesting as they indicate that the average delay value can be either minimized or maximized through a careful choice of spatial location of the two pauses used in the simulation. Indeed, the minimum value of the delay occurs when the two pauses have spatial locations that are very close; however, the times at which the pauses occur are significantly different. In addition, the maximum value of the average delay also occurs when the pauses have spatial locations that are still reasonably close together in terms of their spatial location. Of course, the time durations of the two pauses also influence where the minimum and maximum values occur, and this can be seen in greater detail in [2].

Using the basic insights of the above discussion, it is a simple matter to place a cluster of pauses in space and time so that the average transcription time per polymerase is increased significantly from that of the reference case of no pauses. To illustrate this idea, we place a total of eight pauses in the computational domain and compute the average delay per polymerase for this simulation. Note that the general placement of the cluster in this example is informed by the experimental results in Figure 4 of [3] indicating the area of the DNA template where the transcriptional pauses are most likely to be observed. Using the framework outlined in (4a)-(4c), the background density for the simulation is $z_0 = 0.31$. In order to prescribe the pauses, we use (5) with $M = 8$, and Table I gives a list of the spatial locations of eight pauses (x_i) with their corresponding start times t_i. All the pauses have the same time duration of $d_i = 0.05$ for $i = 1, 2, \ldots 8$, and the simulation was computed on a time interval of $(0, 5.3665)$, a sufficient time for the effects of all pauses to have passed through the termination end of the strand. The resulting average delay per polymerase is 0.0945. For the background density $z_0 = 0.31$, the transcription time per polymerase for the reference case is $1/0.69 \approx 1.4493$, and the average delay value computed for the simulation with eight pauses corresponds to approximately a 13.7% increase in the average transcription time for each RNA polymerase over this limited region of time. Hence, this simple example illustrates how the clustering of pauses can increase the time needed for each RNAp to transcribe the DNA strand thereby reducing the overall amount of mRNA that the gene is likely to be able to produce in its lifespan. As in the previous examples, we have chosen a particular placement of pauses, and an investigation of clusters of pauses located randomly throughout the space and time domain would require a statistical analysis over repeated sampling of the underlying pause distribution. One goal of the current research is to apply sensitivity analysis techniques to the traffic flow model considered here in order to systematically investigate a similar clustering phenomena in this context.

Location	Start Time
$x_1 = -0.450$	$t_1 = 0.1$
$x_2 = -0.300$	$t_2 = 0.3$
$x_3 = -0.375$	$t_3 = 0.5$
$x_4 = -0.225$	$t_4 = 0.6$
$x_5 = -0.300$	$t_5 = 0.8$
$x_6 = -0.150$	$t_6 = 0.8$
$x_7 = -0.225$	$t_7 = 1.0$
$x_8 = -0.150$	$t_8 = 1.2$

TABLE I

TABLE DESCRIBING THE PLACEMENT OF PAUSES IN THE SPACE AND TIME DOMAIN. ALL PAUSES HAVE A DURATION OF 0.05 UNITS OF TIME.

V. CONCLUSIONS

This work presents a parameter study for numerical simulations of a nonlinear PDE model that describes DNA transcription in the presence of short pauses that are located throughout the spatial and time domains. A Discontinuous Galerkin finite element method is used for the numerical simulations because of the algorithm's ability to maintain stability and to accurately capture multiple shocks and rarefaction waves (created by the incorporation of transcriptional pauses) within the solution behavior. In highly transcribed genes such as the *rrn* gene in E.coli under high growth conditions, the simulation of the interaction of high densities of polymerases with frequent pauses requires employment of

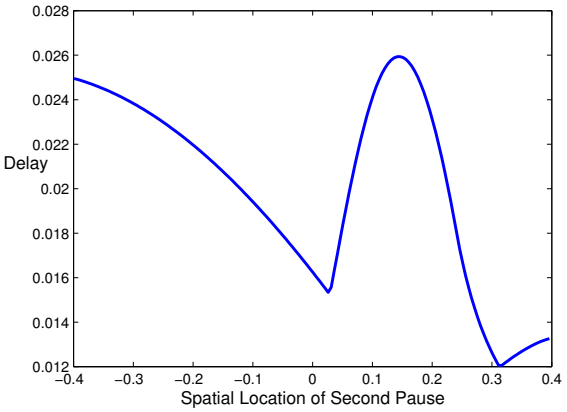

Fig. 4. Average delay per polymerase as a function of the spatial location of the second pause.

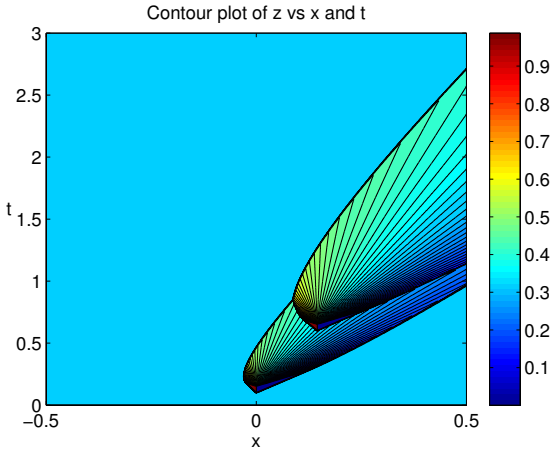

Fig. 5. Contour plot of the density $z(x,t)$ where the location of the second pause corresponds to the local maximum value of the delay.

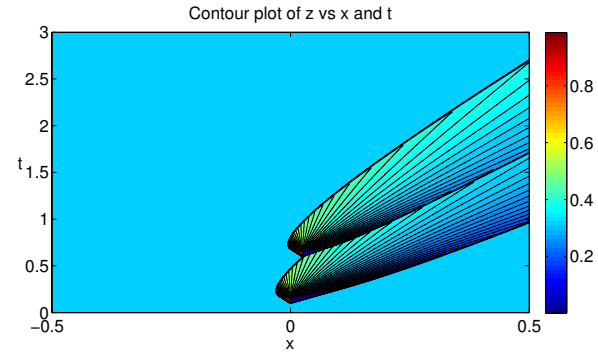

Fig. 6. Contour plot of the density $z(x,t)$ from the parameter study in section IV-B where the location of the second pause corresponds to the first local minimum value of the delay.

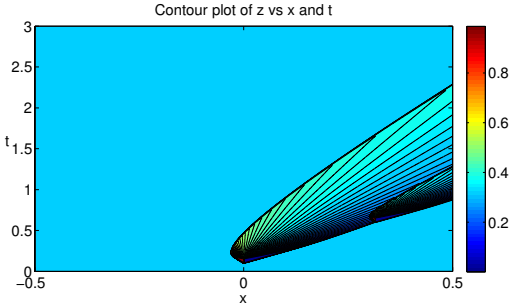

Fig. 7. Contour plot of the density $z(x,t)$ from the parameter study in section IV-B where the location of the second pause corresponds to the second local minimum value of the delay.

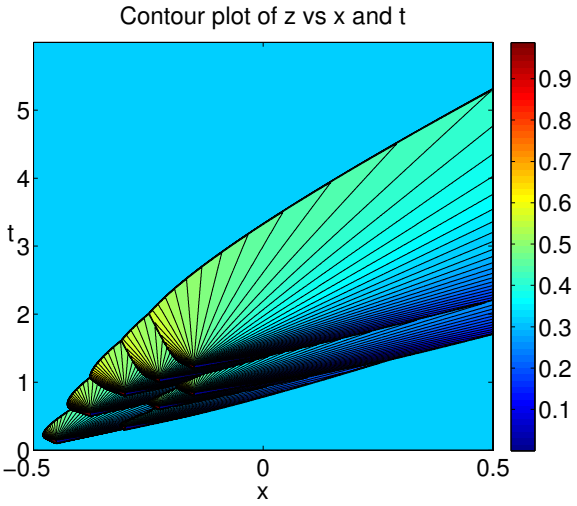

Fig. 8. Contour plot of the PDE simulation corresponding to a cluster of $M = 8$ pauses located in the domain according to the information in Table I and using $z_0 = 0.31$.

highly sophisticated numerical methods that are stable in the presence of multiple discontinuities in the coefficients, as well as solutions of a nonlinear conservation law.

Motivated by the underlying biological problem we calculate the average transcriptional time delay per polymerase as a measure of the effect that the ubiquitous pauses may have on the overall production of ribosomes (and hence proteins) in E. coli. The analysis of the PDE model with many pauses randomly distributed throughout the domain is difficult and will require a statistical approach. Here we take the first steps toward a different approach by considering a parameter study for simple models such as the two pause model described in Sections IV-A and IV-B. We also include an example to illustrate that a cluster of pauses can be constructed that can significantly influence the average transcription time required by an RNAp. We demonstrate that important parameters include the spatial location and the time duration of the pauses as well as the background density of the PDE (z_0 in our notation) which describes the recruitment of RNAP to a particular gene. A development of mathematical models for transcriptional elongation and numerical, statistical and analytical analysis of these models will continue to probe the limits that this process imposes on the growth rate, as well

as other cell functions.

ACKNOWLEDGMENT

D.G.T. thanks the National Science Foundation for their support in funding this work.

REFERENCES

[1] L. Davis, T. Gedeon, J. Gedeon, and J. Thorenson, "A Traffic Flow Model for Bio-polymerization Processes," in Journal of Mathematical Biology, 2013, Print ISSN: 0303-6812, Online ISSN: 1432-1416.

[2] L. Davis, T. Gedeon, and J. Thorenson, "Discontinuous Galerkin Calculations for a Nonlinear PDE Model of DNA Transcription with Short, Transient and Frequent Pausing," In revision for Journal of Computational Mathematics, March 2013.

[3] K. Neuman, E. Abbondanzieri, R. Landick, J. Gelles and S. Block. "Ubiquitous transcriptional pausing is independent of RNA polymerase bactracking" Cell, 115(4):437-447, 2003.

[4] H. Bremer and P. Dennis. Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, in chapter Modulation of chemical composition and other parameters of the cell by growth rate, page 421. ed. Neidhardt EA (Sinauer, Sunderland, MA) 2nd Ed., College Station, Texas, 1996.

[5] C. Condon, S. French, C. Squires, and C. Squires. "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.

[6] Stefan Klumpp and Terence Hwa. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. PNAS, 105(47):18159-164, 2008.

[7] B. Cockburn, "Discontinuous Galerkin methods for convection-dominated problems," in High-order Methods for Computational Physics, volume 9 of Lect. Notes Comput. Sci. Eng., pages 69-224. Springer, Berlin, 1999.

[8] J. Hesthaven and T. Warburton, Nodal Discontinuous Galerkin Methods: Algorithms, analysis, and applications, volume 54 of Texts in Applied Mathematics, Springer, New York, 2008.

[9] S. Gottlieb, C-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. SIAM Review, 43(1):89-112 (electronic), 2001.

[10] R. LeVeque. *Finite Volume Methods for Hyperbolic Problems.* Cambridge University Press, New York, NY, 2002.

[11] D. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. "Unified analysis of discontinuous Galerkin methods for elliptic problems," SIAM J. Numer. Anal., 39(5):1749-1779, 2001/02.

[12] B. Cockburn and C-W. Shu. "Runge-Kutta discontinuous Galerkin methods for convection-dominated problems," J. Sci. Comput., 16(3):173-261, 2001.

[13] R. Landick. Transcriptional pausing without backtracking. *PNAS,* 106(22):8797-8, 2009.

[14] J. Thorenson, Discontinuous Galerkin Finite Element Method for Simulation of a Transcription Process Model Ph.D. dissertation, Dept. of Math. Sci., Montana State Univ., Bozeman, MT, 2013.

[15] M. Lighthill and G. Whitham "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, May 10 1955.

[16] P.Richards, "Shock waves on the highway," Operations Research, 4(1):42-51, 1956.

[17] T. Chou and G. Lakatos, "Clustered Bottlenecks in mRNA Translation and Protein Synthesis," in The American Physical Society, Vol. 93, Num. 19, 2004.