Journal of Computational Mathematics
Vol.32, No.6, 2014, 601–629.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1405-m4370

DISCONTINUOUS GALERKIN CALCULATIONS FOR A NONLINEAR PDE MODEL OF DNA TRANSCRIPTION WITH SHORT, TRANSIENT AND FREQUENT PAUSING*

Lisa Davis Tomáš Gedeon
Department of Mathematical Sciences, Montana State University,
P.O. Box 172400, Bozeman MT 59717
Email: davis@math.montana.edu gedeon@math.montana.edu
Jennifer Thorenson
Department of Mathematics, University of Oregon, Eugene, United States
Email: jthoren2@uoregon.edu

Abstract

A discontinuous Galerkin finite element method is used to approximate solutions to a classical traffic flow PDE. This PDE is used to model the biological process of transcription; the process of transferring genetic information from DNA either to mRNA or to rRNA. The transcription process is punctuated by short, frequent RNAP pauses which are incorporated into the model as traffic lights. These pauses cause a delay in the average transcription process. The DG solution of the nonlinear model is used to calculate the delay and to determine the effect of the pauses on the average transcription time. Numerical error measurements between the DG solution and the true solution (derived by the method of characteristics) are given for a simple model problem. It shows an excellent agreement in a neighborhood away from the shocks as well as $\mathcal{O}(\Delta x)$ convergence for the delay calculation. Preliminary parameter studies indicate that in a system with multiple pauses both the location and time duration of the pauses can significantly affect the average delay experienced by an RNAP.

Mathematics subject classification: 65M60, 35R05, 35Q92, 35L65.
Key words: Discontinuous Galerkin finite element method, Transcription, LWR traffic flow model.

1. 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. As noted in [20], each of these processes is quite complicated and requires the availability of various amino acids as well as assembly of various

* Received March 12, 2013 / Revised version received May 4, 2014 / Accepted May 26, 2014 /
Published online September 3, 2014 /

initiator and termination complexes. The recent paper by Zia and collaborators [20] gives an excellent overview of these processes for the reader that may be unfamiliar with the subject. Moreover, the authors point out that the systematic study of such a process can follow either of two paths. The first is the path of building a complex and detailed model that may provide quantitative information that can be compared with experimental data of a very particular setting. Typically, it is very difficult to mathematically analyze the model or to extract from it any fundamental behaviors associated with the physical phenomena. An alternative approach is to consider a simple model that includes only a few of the fundamental underlying mechanisms of the process but hopes to capture the behavior of the system in some general sense. The disadvantage of such an approach is that these types of models rarely exhibit the kind of fidelity that allows for comparison with experimental data. The focus of this paper is the latter approach where a relatively simplified mathematical model is considered as an attempt to gain insight into the limiting behavior of the protein production process under certain cellular conditions.

On a very basic level both transcription and translation consist of a motion of a complex machine along a one-dimensional strand. Each process 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 elongation within the transcription stage, that is, RNAP motion along the DNA strand. 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 [17] 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 [13]. One type is referred to as backtracking pauses, and the other is called non-backtracking pauses, described in [11]. For the purposes of this work, we are only concerned with the non-backtracking pause case.

The existence of transcriptional pauses brings up the possibility that under conditions of high transcription initiation rates, a paused polymerase can prevent the forward motion of one or more polymerases that follow it, thus creating traffic jams [12]. 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 [1], 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 [6], 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 serves as a companion to that work, and it details the Discontinuous Galerkin finite element methods used for the simulation results of the PDE model used in [6].

We present a study of the behavior of numerical solutions of various formulations of the nonlinear model, and these results are used as a means to analyzing the fundamental biological questions related to the transcription process. 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 model problems of interest. This allows us to validate our simulation results prior to tackling the more relevant biologically motivated model formulations. 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.

Section 2 begins with a 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 conservation law form is considered. Solutions of the PDE models considered here exhibit discontinuities for intervals of time 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 sweep through the domain once the transcriptional pause has ended. Hence, the numerical scheme is constructed with the intent of accurately capturing this behavior. For certain formulations of the model problem, errors in the numerical approximations can be measured, and we observe the correct order of convergence away from the spatial location of the discontinuities 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. Once the numerical framework is established, we move to a description of the mathematical models.

In Section 3, we first present a generic version of the nonlinear model problem, which is also stated in the form of a conservation law. The subsections that follow survey several particular cases of that generic form. We relate these particular formulations to the biological phenomena when possible. A closed form analytical solution is constructed for the simplest cases related to one pause location at the center of the DNA strand. The analytical solutions and the numerical computations demonstrate excellent agreement for the simple models. The approximation error associated with the numerical scheme is given, and we discuss convergence properties of the computational scheme.

Once the numerical simulations are verified for the simple model problems, Section 4 details the measure that we use to analyze the throughput of the transcription process under the condition that transcriptional pausing occurs. First, the *average delay per polymerase* is defined in terms of the model variables (density and flux). We then describe how the DG simulation results are used to estimate this value. Numerical errors and convergence rates are assessed for one of the model problems considered in Section 3. Once this proof-of-concept is established, Section 4.3 considers a more biologically realistic case where multiple pauses are introduced into the system and an average delay per polymerase is computed. A discussion of the computational challenges is included, and the complexities of those calculations lead us to Sections 5.1 and 5.2 which employ parameter studies. These subsections illustrate how pause location and pause duration times can significantly affect the average delay values. In turn, the behavior of this average delay function gives us insight into the effects that the transcriptional pausing mechanism may have on the total protein production of the cell.

2. 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 [14] with the geometric flexibility and high-order accuracy of the finite element method [2, 10]. 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 along with computational advantages and some disadvantages.

2.1. 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

$$z_t + [f(t,x,z)]_x = 0, \quad x \in [L,R], \quad t \geq 0, \tag{2.1}$$

where $f(t,x,z)$ is the 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 [2–4, 10] and the references contained therein.

Let the function z_h denote the DG approximation of the solution to Eq. (2.1) 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 (2.1) on the k-th element of the mesh obtained by using a grid of size determined by the parameter h. 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 [10] and references therein. Matlab code obtained from the website associated with [10] 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 needed. In the development of finite volume methods, slope limiters were derived to guarantee that a first order forward Euler scheme was stable when applied to the semi-discrete system of ODEs for a nonlinear hyperbolic PDE. But to gain higher accuracy than first order in time, a class of Runge-Kutta methods were developed to be total variation diminishing and therefore stable whenever an Euler method was also TVD as discussed in [8]. Since slope limiters are needed to guarantee stability of an Euler method, they are also needed for stability of the RK method.

Consider a standard explicit Runge-Kutta method with s stages denoted by

$$\begin{aligned}
&z^{(0)} = z^n,\\
&z^{(i)} = \sum_{m=0}^{i-1} \alpha_{i,m} z^{(m)} + \beta_{i,m}\Delta t L(z^{(m)}, t^n + \gamma_m \Delta t) \quad \text{for } \ i = 1, \cdots, s,\\
&z^{n+1} = z^{(s)},
\end{aligned}$$

where

$$\Delta t \le c\Delta t_{FE}, \quad c = \min_{i,k} \frac{\alpha_{i,k}}{\beta_{i,k}}$$

and Δt_{FE} is the time step from the forward Euler scheme.

As shown in [8], if

$$\alpha_{i,m} \ge 0 \qquad \text{and} \qquad \sum_{m=0}^{i-1} \alpha_{i,m} = 1,$$

then the RK stages are convex combinations of forward Euler steps. Therefore if a forward Euler time integration is TVD, each stage of the RK integration will also be TVD and so will the resulting RK method.

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. It is based on a piecewise-linear reconstruction from finite volume methodology. Suppose the local solution on element D^k has the form

$$z_h^k = \bar{z}_h^k + (x - x_0^k)(z_h^k)_x,$$

where x_0^k is the center of D^k and $\bar{z}_h^k$ is the average value over the element. Then the slope limited version of the local solution is

$$\prod z_h^k(x) = \bar{z}_h^k + (x - x_0^k) m \left((z_h^k)_x, \frac{\bar{z}_h^{k+1} - \bar{z}_h^k}{h}, \frac{\bar{z}_h^k - \bar{z}_h^{k-1}}{h} \right), \tag{2.2}$$

where the `minmod` is defined to be

$$m(a_1,\cdots,a_n)=\begin{cases} s\ \min_{1\le i\le n}|a_i|, & |s|=1\\ 0 & |s|\neq 1\end{cases}\qquad s=\frac{1}{n}\sum_{i=1}^{n}\operatorname{sign}(a_i).$$

The `minmod` slope limiter compares an approximation to the spatial derivative over the element with difference quotients using the two neighboring elements. If these slopes all have the same sign, then the slope of the local solution over the element is chosen to be the smallest in absolute value. If the slopes do not have the same sign, then an oscillation is detected, the slope is set to be zero, and the local solution is the average value over the element. 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. It is shown in the following sections that this issue causes very little concerns for our simulations; however, it must be taken into account when convergence studies are employed. All calculations in the current work are obtained using the following optimal third order Runga-Kutta method with slope limiting. The scheme is optimal in the sense that the time step can be chosen to be as large as the time step for the forward Euler method. Numerical implementation of the algorithm is as follows. First, we apply the slope limiter to the initial condition.

$$z(x,0)=z_0(x)\quad\Rightarrow\quad z^0=\prod z_0(x)$$

where $\prod$ is the minmod slope limiter in (2.2). Then for $n=0,1,2,\cdots,m$, successive iterations of the slope-limited TVD Runge Kutta algorithm obtain a numerical approximation z^n as follows.

$$\begin{aligned} z^{(1)}&=\prod[z^n+\Delta tL(z^n,t^n)]\\ z^{(2)}&=\prod[\tfrac{3}{4}z^n+\tfrac{1}{4}z^{(1)}+\tfrac{1}{4}\Delta tL(z^{(1)},t^n+\Delta t)]\\ z^{n+1}&=\prod[\tfrac{1}{3}z^n+\tfrac{2}{3}z^{(2)}+\tfrac{2}{3}\Delta tL(z^{(2)},t^n+\tfrac{1}{2}\Delta t)]\end{aligned}\tag{2.3}$$

where 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 in the spatial domain. Using this DG formulation for a scalar nonlinear conservation law, the next section outlines its implementation for the nonlinear PDE used to mathematically describe the biological process of transcription.

3. Nonlinear Models

Here we discuss the general framework common to all of the models considered in this paper, and with this framework established, various model simulations and convergence results are presented. We examine a nonlinear macroscopic PDE model that was originally considered as a continuum model for traffic flow, see [15,18]. 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. This equation takes the form

$$\begin{aligned}&z_T + [\bar{\beta}(X,T)(1-z)z]_X = 0 \quad X \in (-L/2, L/2), \quad T > 0\\ &z(X,0) = z_0\\ &z(-L/2,T) = z_0\end{aligned}$$

where T is time measured in seconds, X is a continuum version of the length of a DNA strand that would be measured in units of the number of nucleotides in a discrete setting, and L is the length of the DNA strand in terms of those units. The function $z(X,T)$ represents the density of RNAP on the DNA strand as a function of space and time. The location and time duration of the transcriptional pauses are specified through the definition of the coefficient $\bar{\beta}(X,T)$, within the flux function, in a piecewise manner simply as

$$\bar{\beta}(X,T) = \begin{cases} 0 & \text{pause} \\ v & \text{no pause} \end{cases} = v \begin{cases} 0 & \text{pause} \\ 1 & \text{no pause} \end{cases} = v\beta(X,T), \tag{3.1}$$

where v is the (constant) maximum velocity. (A more detailed description of $\bar{\beta}(X,T)$ will be specified later by a spatial pause location, beginning pause time and pause duration.)

Time and space are rescaled using the dimensionless variables

$$t = \frac{vT}{L} \quad \text{and} \quad x = \frac{X}{L}.$$

By the Chain Rule, the original PDE is equivalent to

$$\frac{v}{L} z_t + \frac{1}{L}[v\beta(x,t)(1-z)z]_x = 0, \quad x \in (-0.5, 0.5), \quad t > 0.$$

Therefore, the dimensionless PDE model is

$$z_t + [\beta(x,t)(1-z)z]_x = 0, \quad x \in (-0.5, 0.5), \quad t > 0, \tag{3.2a}$$

$$z(x,0) = z_0, \tag{3.2b}$$

$$z(-0.5,t) = z_0. \tag{3.2c}$$

We use the flux function to impose transcriptional pauses. In particular, $\beta(x,t)$ is determined by the nature of the transcriptional pauses and is defined in the following manner. For $i = 1, 2, \ldots M$, 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.

$$\beta(x,t) = \begin{cases} 0 & x = x_i \text{ and } t_i < t < t_i + d_i, \text{ for } i = 1, 2, \ldots M, \\ 1 & \text{otherwise.} \end{cases} \tag{3.3}$$

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. In Section 4.3, we discuss computational results for a very large value of M that is biologically relevant. Prior to presentation of those results, we present a survey of numerical results for simplified models which are used to validate the accuracy and convergence rates of the algorithms constructed for simulation of the more biologically relevant mathematical models.

For details related to the biological literature, the interested reader is referred to [6] where the authors discuss several aspects of the motivating questions from the biological system and give computational results for the general model given above. In that work, the focus is the justification of the nonlinear model as a viable mathematical model and the numerical results are described in terms of their implications for the biological mechanisms related to physical phenomena of transcriptional pausing. This work focuses on the validation of the numerical implementation and the resulting computational simulations. In the following sections, we include a detailed computational study of the accuracy of the simulations. Beginning with two cases where one can construct an analytical expression for a solution, we give a careful study of the convergence properties of the simulation algorithm. Once these results are established, we include other computational examples that are relevant to the biological model but where an algebraic solution is not available.

3.1. RNAP elongation after a pause

In this section, we examine the performance of the DG method on the nonlinear model using a smooth flux function but a piecewise constant initial condition modeling only the portion of the transcription process that occurs after one paused nucleotide releases the lead RNAP and elongation is allowed to take place without any other pauses occurring. We consider this case purely from the standpoint of verification of the simulations as there is a true solution that we can construct explicitly. This very simple case serves as a pre-cursor to model problems that are more relevant to the biological questions we seek to answer and simultaneously less amenable to any existing convergence results. We numerically solve

$$z_t + [(1-z)z]_x = 0 \quad x \in (-0.5, 0.5), \quad t > 0, \tag{3.4a}$$

$$z(x,0) = \begin{cases} 1 & x < 0, \\ 0 & x > 0, \end{cases} \tag{3.4b}$$

$$z(-0.5,t) = 1. \tag{3.4c}$$

In terms of its application to ribosomal transcription, this model describes an idealized case where a pause with a long duration has occurred in the middle of the DNA strand ($x = 0$), the density in front of the pause has dropped to 0, and the trailing RNAPs have backed up behind the first stopped polymerase. Using the method of characteristics, one can find the closed form solution to (3.4a)-(3.4c) to be the following function.

$$z(x,t) = \begin{cases} 1 & \text{if } x < -t, \\ \frac{1}{2}\left(1 - \frac{x}{t}\right) & \text{if } -t \le x \le t, \\ 0 & \text{if } x > t. \end{cases}$$

Note that the solution is piecewise defined and is determined by a linear function in one region of the domain and taking on a constant value in the other two regions; hence, we choose to use a linear basis for the spatial discretization of the approximation scheme. Note that Theorem 4.7 in Chapter 4 of [10] implies that for any $t > 0$, the L^2 norm of the error of the approximation is bounded by a constant multiple of Δx. Combining this spatial approximation with the explicit TVD Runge Kutta time integration and the use of the slope limiter suggest that $\mathcal{O}(\Delta x)$ is the best order of convergence that one might expect.

Using a linear basis, the L^2 error at the final time T is computed for a range of spatial discretizations from $\mathcal{K} = 40$ to $\mathcal{K} = 1280$ elements. In Table 3.1, the error is computed over the spatial domain $x \in (-0.5, 0.5)$ at the final time $T = 0.7$, and note that the convergence rate of $\mathcal{O}(\Delta x)$ is as expected. We also remark that the boundary condition imposed in (3.4c) is inconsistent with the true solution to (3.4a)-(3.4b) for $t > 0.5$. That is, at the time $t = 0.5$, the characteristics emanating from the shock at $x = 0$ first intersect the left end of the domain, $x = -0.5$, and the true solution values that are determined by the characteristic lines (as opposed to using those values prescribed by the boundary condition in (3.4c)) are used in the error calculation in order to recover the prescribed error convergence rate. We should remark here that if the boundary condition in Eq. (3.4c) is used in that computation, the DG approximation seems to isolate the error in the boundary condition to that element containing the boundary node $x = -0.5$. Hence, even when the prescribed boundary condition violates the conditions required by the characteristics, the DG approximation still exhibits reasonable convergence as the grid is refined; however, the order of the convergence rate is slightly below $\mathcal{O}(\Delta x)$. The following section contains a model that is more relevant to the mathematical model of interest but has been simplified to a degree such that one can write down a closed-form solution.

Table 3.1: Table of error calculations for the model in Section 3.1 - transcription after a pause ends. The L^2-error is calculated at the final time $T = 0.7$ showing that the order of convergence is $\mathcal{O}(\Delta x)$.

$\mathcal{K}$	$L^2(-0.5, 0.5)$	order
40	0.0091749	
80	0.0045517	1.0112837
160	0.0022671	1.0055712
320	0.0011315	1.0026353
640	0.0005653	1.0012029
1280	0.0002828	0.9989672

3.2. Transcription with one pause in a parameterized interval of time

In this section, we consider the model in (3.2a)-(3.2c) with the configuration that one nucleotide in the center of the DNA strand corresponds to a pause location and the pause occurs during some parameterized time interval denoted by $\zeta < t < \xi$. The flow velocity using the piecewise constant function is specified as

$$\beta(x,t) = \begin{cases} 0, & \text{if } x = 0 \quad \text{and} \quad \zeta < t < \xi \\ 1, & \text{otherwise.} \end{cases} \tag{3.5}$$

It represents (a dimensionless version of) 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 $dt = \xi - \zeta$. During this time period, the velocity function is set to 0, and no elongation is permitted at this spatial location. The location of the pause or the so-called bottleneck nucleotide is fixed at $x = 0$ for ease of computation.

Using the method of characteristics, one can develop a closed-form solution to (3.2a)-(3.2c) with (3.5), see [9] and [16] for example. For brevity, we omit the details of this calculation and refer the reader to the Appendix of this paper. One can define specific times t_1 and t_2 so that 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. With $dt = \xi - \zeta$, $t_1 = \frac{dt}{1-z_0}$ and $t_2 = \frac{dt}{z_0}$ and $z_0 = 0.31$, the closed-form solution to (3.2a)-(3.2c) using (3.5) is given by

$$z(x,t) = \begin{cases} 0.31 & t < \zeta, \\ 0.31 & x < -0.31(t-\zeta); \quad t < t_1 + \zeta, \\ 1 & -0.31(t-\zeta) < x < 0; \quad t < \xi, \\ 1 & -0.31(t-\zeta) < x < -(t-\xi); \quad \xi < t < t_1 + \zeta, \\ \dfrac{1}{2} - \dfrac{x}{2(t-\xi)} & -(t-\xi) < x < 0; \quad \xi < t < t_1 + \zeta, \\ 0.31 & x < 0.38(t-\xi) - 2\left[\frac{3.1(t-\xi)}{69}\right]^{1/2}(0.69); \quad t \geq t_1 + \zeta, \\ \dfrac{1}{2} - \dfrac{x}{2(t-\xi)} & 0.38(t-\xi) - 2\left[\frac{3.1(t-\xi)}{69}\right]^{1/2}(0.69) < x < 0; \quad t \geq t_1 + \zeta, \\ 0 & 0 < x < (0.69)(t-\zeta); \quad t < \xi, \\ \dfrac{1}{2} - \dfrac{x}{2(t-\xi)} & 0 < x < t - \xi; \quad \xi < t < t_2 + \zeta, \\ 0 & t - \xi < x < (0.69)(t-\zeta); \quad \xi < t < t_2 + \zeta, \\ 0.31 & (0.69)(t-\zeta) < x; \quad t < t_2 + \zeta, \\ \dfrac{1}{2} - \dfrac{x}{2(t-\xi)} & 0 < x < 0.38(t-\xi) + 0.62\left[\left(\frac{69}{31}\right)(t-\xi)(0.1)\right]^{1/2}; \quad t \geq t_2 + \zeta, \\ 0.31 & 0.38(t-\xi) + 0.62\left[\left(\frac{69}{31}\right)(t-\xi)(0.1)\right]^{1/2} < x; \quad t \geq t_2 + \zeta. \end{cases} \tag{3.6}$$

Note that a discussion similar to the one from the Appendix can be found in [16] with the exception that the explicit form of the solution is not given in that text and the paper models a red-green traffic light using a discontinuous initial condition while we use a discontinuous flux coefficient. The parameterized pause coefficient allows one to consider multiple spatial locations for the pauses, different intervals of time as well as different durations of time for the model, thereby, making it useful for comparison to the actual biological applications that are being considered. The true solution (3.6) is used to validate numerical simulations of the PDE and to study the delays experienced by polymerases in the presence of a single pause nucleotide.

Also note that similar calculations for a particular instance of the model in (3.2a)-(3.2c) with (3.5) using DG have appeared in print, see [19]. The results in this paper differ from those in several ways. In [19], 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 biopolymerization. The following section presents a set of computational results using the DG scheme for the model given in this section.

3.2.1. Error calculations and convergence results

Fig. 3.1 gives a sample of a contour plot for the model discussed above with the choices of $z_0 = 0.31$ where $\beta(x,t)$ incorporates a pause at $x = 0$, $\zeta = 0.2$ and $\xi = 0.3$ from Eq. (3.5).

Fig. 3.2a shows a plot of the numerical approximation of density along with the true solution, given in Eq. (3.6), at the final time $T = 0.7$ using $z_0 = 0.31$ and 160 elements. The pointwise error between the DG approximation and the solution at that final time is also shown in Fig. 3.2b. We note that for the graph in Fig. 3.2b, the pointwise error spikes in the region of the domain where the shock is located for that particular value of t. Although the error is larger in the region between the two shocks, it is still within an acceptable tolerance for the value of $\Delta x = 6.25 * 10^{-3}$ using 160 elements if one might expect $\mathcal{O}(\Delta x)$ convergence. In order to investigate convergence properties, we use a linear basis with the DG scheme discussed in Section 2.1, and the L^2 error at the final time $T = 0.7$ is computed for a range of spatial discretizations. Due to a loss of accuracy near the shocks, the error is computed over the spatial region satisfying $|x - x_s| \geq 0.01$ where x_s is the location of the shock at the final time, see [3] for similar error tabulations. Table 3.2 gives the results of the error calculations and shows a decrease in the error outside a small neighborhood of the spatial domain that contains the shock; however, the order of convergence does not indicate any clear pattern.

Table 3.2: Error calculations at the final time $T = 0.7$ where one pause is located at $x = 0$ during the time interval $0.2 < t < 0.3$. Ω is a subset of the domain defined by $\Omega = \{x \in (-0.5, 0.5) : |x - x_s| \geq 0.01\}$ where x_s is the location of a shock at T.

$\mathcal{K}$	$L^2(\Omega)$	order
40	0.0191218	
80	0.0070542	1.4386778
160	0.0013742	2.3598635
320	0.0003147	2.1264354
640	0.0001769	0.8309569
1280	0.0000650	1.4438834

In Fig. 3.3a, the density at the right boundary $x = 0.5$ (corresponding to the termination site of the DNA strand) is plotted as a function of the time t. Observe that the DG calculation shows good agreement with the true solution of the PDE with the exception of small errors in the region around the discontinuities in the density and flux functions. Moreover, the discontinuities are captured by the DG approach without the introduction of any non-physical oscillations around the discontinuities. Numerical error in the DG approximation at the termination site $x = 0.5$ is particularly important to the underlying biological model because the flux at the termination site is measured (over time) in order to calculate the number of polymerases that complete the transcription process (thereby producing ribosomes) in the model simulation. The formulation of this delay calculation is now described in the following section.

4. 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 [6, 12, 13, 17]. One goal of the current research is to use numerical simulations of the PDE model described in Section 3 (with various selections of pause data) 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

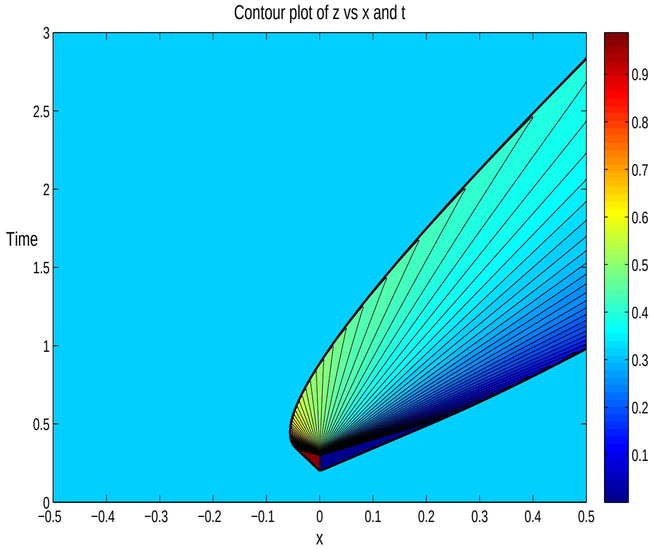
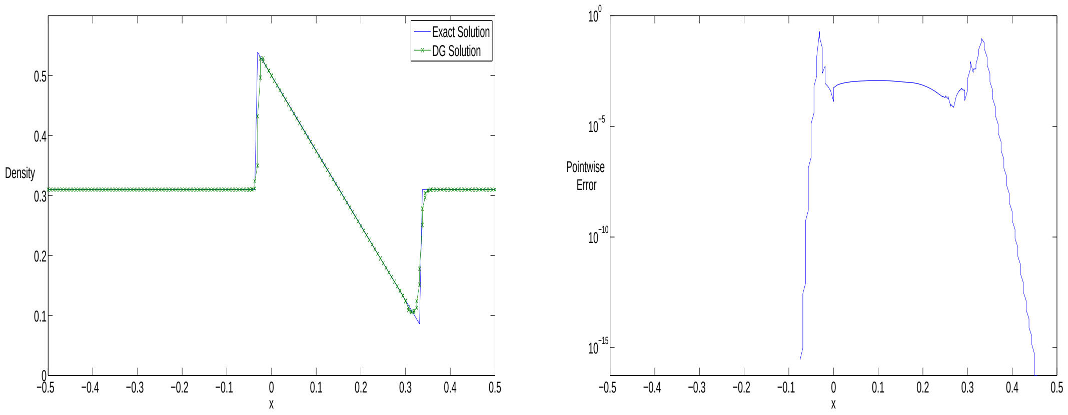


Fig. 3.1.: A contour plot of solution $z(x,t)$ with $z_0 = 0.31$ where $\beta(x,t)$ incorporates a pause at $x = 0$, $\zeta = 0.2$ and $\xi = 0.3$ from Eq. (3.5). The DG solution was simulated using a linear basis and 640 elements.



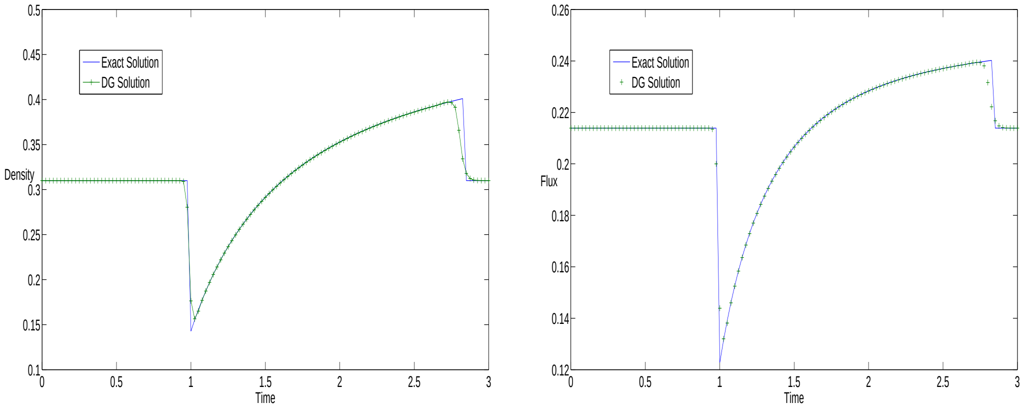
(a) The dashed (green) curve is the graph of the DG simulation of the PDE model in (3.2a). The solid (blue) curve represents the true solution as given in (3.6).

(b) Pointwise Error

Fig. 3.2.: Solution and Pointwise Error at $T = 0.7$ using 160 elements.

pauses, and a sample calculation of this delay value is given for the setting of multiple pauses at the end of the discussion.

Using a PDE model, the information that we compute is a density of RNAPs over the domain of interest. We do not compute trajectories of individual RNAPs. However, one can measure the amount of RNAPs that have successfully transcribed the strand over a chosen



(a) Density at $x = 0.5$.

(b) Flux at $x = 0.5$.

Fig. 3.3.: DG and true solution of the nonlinear PDE at $x = 0.5$ where a pause is located at $x = 0$ during times $0.2 < t < 0.3$ using 160 elements. The solid (blue) curve is the true solution and the dashed green curve is the solution using a DG simulation.

interval of time. Denoting the flux function from (3.2a) as $f(z) = \beta(x,t)(1-z)z$, the number of polymerases that have arrived at the termination site over the time interval $[T_0, t]$ is given by

$$N(t) = \int_{T_0}^{t} f(z(0.5, y))dy, \tag{4.1}$$

where $z(0.5, t)$ is the solution of (3.2a)-(3.2c) and represents density at the termination end of the strand at time t. Using this information, one can compute time delays by comparing the outputs of two versions of the PDE model.

First, we introduce the *reference case* of the model. The *reference case* characterizes the ideal situation in which there is no pause on the DNA strand, and the initial and boundary conditions are specified with a certain constant *background* density of z_0. The PDE model is then posed as Eqs. (3.2a)-(3.2c) where we set $\beta(x,t) = 1$ for all x and t. Note that the constant density $z(x,t) = z_0$ is the solution of the model equation for this reference case; therefore, the reference flux 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 for the reference case with a constant density, constant flux and no transcriptional pauses, Eq. (4.1) implies that the number of polymerases that have arrived at the termination site, $x = 0.5$, in the time interval (T_0, t) is

$$N_R(t) = \int_{T_0}^{t} f_R dy = (t - T_0) f_R. \tag{4.2}$$

This gives us a mechanism to measure the total amount of RNAPs that have transcribed the DNA strand at any given time $t > T_0$ for the ideal case of no pauses.

Next we describe how we compute the effect of transcriptional pauses on the average delay per polymerase. Although the delay calculation is described in terms of the aforementioned model problem, the process is the same for the more general situation where one allows multiple transcriptional pauses to occur. For the case of the model problem with one pause described in Section 3.2, we have constructed the solution of (3.2a)-(3.2c) using (3.5) which incorporates one pause in the middle of the DNA strand. One can use that expression in (3.6) and the corresponding flux $f(z(0.5,t),t)$ to evaluate the corresponding integral in the expression (4.1) to construct $N(t)$ which quantifies the amount of RNAPs that have transcribed the DNA strand at any given time $t > T_0$ for the single pause model.

Then for any $t > T_0$, we calculate the amount of time required for the polymerases transcribing the one pause model to reach the right boundary, and we seek to 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, define the function $s(t)$ so that for each $t > T_0$, the function $s(t)$ is the instant of time for which

$$\int_{T_0}^{s(t)} f(z(0.5,y))dy = N_R(t). \tag{4.3}$$

Note that the function $z(0.5,y)$ in the above equation refers to the density calculated using the one pause model. (In the more general case, $z(0.5,y)$ would refer to the density using a model with multiple pauses imposed.)

Now for each $t > T_0$ in the one pause model, suppose that an amount of RNAP arrives at the termination end of the strand at the time $s(t)$. That amount of RNAP experiences a transcription time delay of $s(t)-t$ (the time it actually arrives at the termination end minus the time it would have arrived at the right end if there were no pauses). Hence, the total amount of delay experienced by all RNAPs arriving over the time interval $[T_0,T_1]$ is given by the following integral

$$\int_{T_0}^{T_1} f(z(0.5,s(t)))[s(t)-t]dt.$$

The average delay per polymerase over an interval $[T_0,T_1]$ is then calculated as

$$D(T_0,T_1) = \frac{1}{\displaystyle\int_{T_0}^{T_1} f(z(0.5,s(t)))dt} \int_{T_0}^{T_1} f(z(0.5,s(t)))[s(t)-t]dt. \tag{4.4}$$

The integral in the denominator is the amount of RNAPs that terminated during the time interval $[s(T_0),s(T_1)]$. In the case of the one pause model, we choose T_0 to be an arbitrary time before the first of the delayed polymerases reached the right boundary and T_1 to be a time after the last of the polymerases affected by the pause has reached the termination site. In this way, we ensure that the delay calculation includes all RNAPs that could possibly be affected by the pause.

Here we note that it is only for the case of the model problem with a single pause that we have an analytical solution that we can use to verify and validate our computational scheme. We will report those results in the following subsections, but first we include the following subsection to describe how all of the ideas presented above get translated into the discrete computational scheme using DG.

4.1. Discretizing the delay calculations

In order to be precise in our validation of the computational results, we briefly introduce notation to clarify how the DG calculations are used to approximate the delay calculations for the PDE model with one pause, and we stipulate that the generalization to a model with more transcriptional pauses follows the same form.

For the system governed by the PDE model and simulated using the DG scheme, the number of polymerases that have arrived at the termination site in the time measured from T_0 to the time t is approximated by

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

where z_h represents the DG solution of (3.2a)-(3.2c) with (3.5). Likewise, we define the discretized function $s_h(t)$ so that for each $t > T_0$, the function $s_h(t)$ is the approximate instant of time for which

$$\int_{T_0}^{s_h(t)} f(z_h(0.5, y))dy = N_R(t). \tag{4.5}$$

The notation s_h represents the fact that the integral is calculated using the DG solution $z_h(0.5, y)$ and by using a numerical quadrature rule to approximate the integral. For the right side of (4.5), the $N_R(t)$ calculation is done exactly. The estimated average delay per polymerase over an interval $[T_0, T_1]$ is then calculated to be

$$D_G(T_0, T_1) = \frac{1}{\displaystyle\int_{T_0}^{T_1} f(z_h(0.5, s_h(t)))dt} \int_{T_0}^{T_1} f(z_h(0.5, s_h(t)))[s_h(t) - t]dt, \tag{4.6}$$

where all quantities that must be numerically approximated are indicated with a form of subscript referring to the mesh size. Note that 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 simple 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.

4.2. Delay calculation for the one pause model

As a means of validating the computational scheme and as an illustrative example, we give the following results. Using the DG solution to $(3.2a) - (3.2c)$ with $z_0 = 0.31$ and

$$\beta(x, t) = \begin{cases} 0, & \text{if } x = 0 \quad \text{and} \quad 0.2 < t < 0.3, \\ 1, & \text{otherwise}, \end{cases} \tag{4.7}$$

the average delay over the time interval $(0, 2.84)$ is computed with the time step size $\Delta t = 0.0001$ for a range of spatial discretizations. Using the true solution (3.6) at $x = 0.5$ to compute the flux function and $s(t)$ from (4.3), the true value of the delay is approximately $D = 0.0422086$. Table 4.1 compares the average delay using the DG solution and that of the true solution showing that the order of convergence is $\mathcal{O}(\Delta x)$.

Now that we have demonstrated convergence of simulations for both the DG solution and the delay calculation in the case of a single pause, we briefly discuss attempts to introduce multiple pauses into the system. In the traffic flow transcription model, multiple pauses are incorporated into the PDE as discontinuities in the flux coefficient to mimic cases where there

Table 4.1: The average delay, D_G for the model in (3.2a)–(3.2c) with (4.7) computed using the DG solution with a variety of spatial grids. The error in the third column is computed by comparing the average delay computed using the DG computation to that of the delay, D as discussed above, computed using the true solution of the PDE.

$\mathcal{K}$	D_G	$\|D - D_G\|$	order
40	0.0316603	0.0105483	
80	0.0374607	0.004748	1.1516275
160	0.0399146	0.002294	1.0494425
320	0.0410612	0.0011474	0.9994793
640	0.0416378	0.0005709	1.0071947
1280	0.0419272	0.0002814	1.0203922

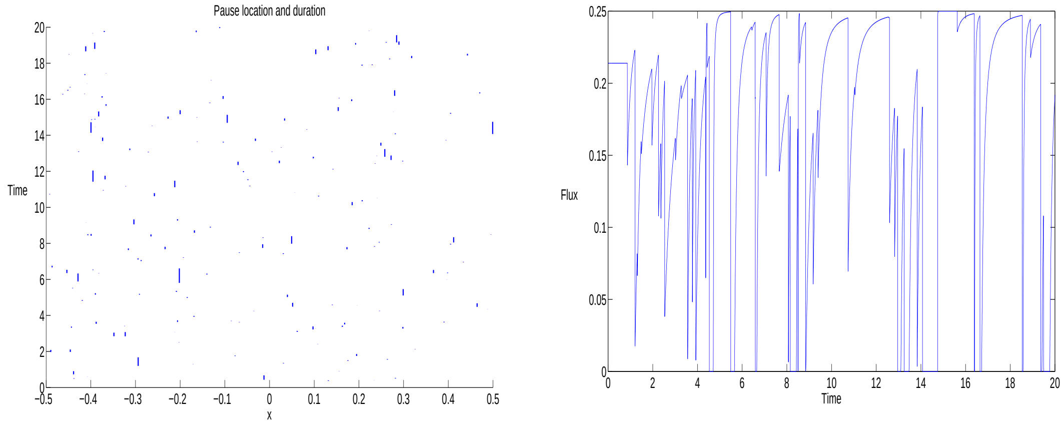
are known experimental results. This is meant to serve as a motivation for considering the parameter studies in the following section.

4.3. Computational results for a biologically relevant model

In [17], single molecule observation using optical traps indicate that the elongating polymerase transcribes rapidly but frequently pauses. Here we construct a mathematical framework that imposes a set of pauses to model that situation. Consider the model equation given in (3.2a)-(3.2c) using the general form (3.3) with a large number of pauses implemented as follows. The spatial domain $[-0.5, 0.5]$ is partitioned into N 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 N, 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 a double exponential distribution. Using experimental data from [17], 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 overall mean length of a pause is then $\tau = 3.12$ seconds. The numerical simulations are constructed to be consistent with [7, 12, 17] where, on average, a polymerase encounters 6 pauses per minute of elongation. Using $v = 200$ in (3.1), 5 minutes of transcription time is equivalent to $t = 300 * v/5450 = 11$ units of dimensionless time where the length of the DNA strand is $L = 5450$ nucleotides. The PDE solution was simulated until a final time of $T = 20$ to ensure that $s_h(t)$ was computed on the dimensionless time interval (1, 12) corresponding to a time interval of length 5 minutes. We begin the delay calculation at $T_0 = 1$ in order to guarantee that the RNAPs arriving at or after $T_0 = 1$ have transcribed either all or most of the total length of the domain and on average, have experienced pauses. Because the initial condition is $z_0 = 0.31$, there will be RNAP that are distributed along the DNA strand at $t = 0$ and those RNAPs will not be as likely to experience a transcriptional pause as those that initiate on or after time $t = 0$. Using 176 pauses uniformly distributed in space and exponentially distributed in time as shown in Fig. 4.1a, the average delay per polymerase for this particular realization of pauses on the interval (1, 12) is 41 seconds. The stochastic nature of the pauses creates a flux function (at the termination site) which is highly oscillatory, see Fig. 4.1b. Fig. 4.1a is included so that the reader may visualize the location and time durations of pauses that result in such a flux function. Even though the flux function

$f(z_h(0.5,t))$ is oscillatory, it's easily integrable using a composite rule with a fine grid of time values as we have used here.

Note that the above result of an average delay per polymerase refers to an average value over time for one particular realization of pauses adhering to the statistical distribution described above. However, we note that an average delay per polymerase for one single realization of pauses (i.e. one draw from a random variable) is certainly insufficient to make any conclusions about overall behavior of such a system. A preliminary result is reported in [6] where multiple realizations of the pauses are used for model simulations, and average delays are reported over a range of values of instantaneous transcription rates. In that case, the average delay calculations exhibit a high amount of variability for the multiple realizations. Even formulating any conclusions about the influence that the location and duration times of pauses might have on the average delay per polymerase is clearly very difficult for such a general case with a large number of pauses located randomly within the domain. The stochastic nature of the pause distribution will require the use of statistical techniques in order for a systematic analysis to lead to any useful qualitative conclusions, and those efforts are currently underway. In the meantime, we make use of the PDE model simulations in order to numerically investigate the system model in the presence of a small number of pauses where the locations and durations become parameters of great interest.



(a) Space and time plot of 176 pauses. The blue indicates the points for which $\beta(x,t)$ is zero.

(b) Flux at $x = 0.5$ using the DG solution incorporating 176 pauses in the discontinuous coefficient.

Fig. 4.1.: Multiple pause example producing an oscillatory flux function at the termination end of the DNA strand.

5. Parameter Studies for Models with Multiple Pauses

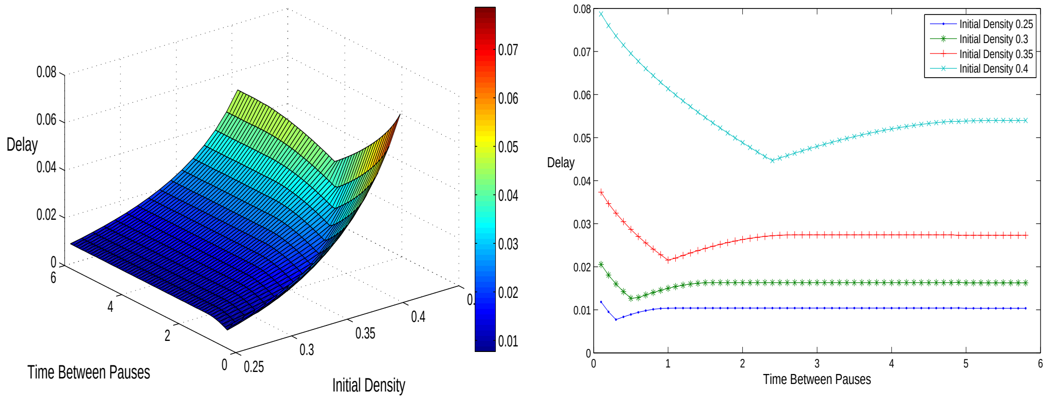
The previous section indicates that a statistical analysis is needed for the biologically relevant situation where a large number of pauses is included in the mathematical model. However, as a preliminary quantitative study where one can make use of the PDE model, we demonstrate that even for fairly simple cases, the delay function can display complicated behavior. We consider a variety of parameterizations of the spatial locations and time durations of the pauses. In [6], the authors show that for the one-pause model discussed in Section 3.2, 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 crossing 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.

5.1. 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

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

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 Eq. (3.2b). 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 RNAPs simultaneously, while in low growth conditions this number can be close to zero.



(a) Surface plot of the average delay per polymerase for a range of initial density values (3.2b) and time between the two pauses.

(b) Cross sections of the average delay plot for several initial density values z_0 in (3.2b) and time between the two pauses.

Fig. 5.1.: Results of parameter study in Section 5.1.

Fig. 5.1a shows that for each choice of background density z_0, there is a minimum delay value (Fig. 5.1b), 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 units of dimensionless time. Focusing on the case of the highest background density of $z_0 = 0.4$, the transcription time for the reference case in terms of dimensionless time corresponds to $1/(1 - z_0) \approx 1.667$. Then an average delay of less than or equal to 0.08 corresponds to a increased transcription time that ranges between 0 and slightly less than 5%

when compared to the transcription time for the reference case. Fig. 5.2 includes the contour plot for the simulation corresponding to the minimum delay value using an initial density of $z_0 = 0.3$. For that case, the maximum amount of delay is around 0.02 units of dimensionless time which corresponds to a worst case delay of only about 1.5% of the dimensionless transcription time. Upon careful mathematical 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 affected by the first pause are not affected by the second pause in such a situation. Moreover, in the nonlinear model, the elongation velocity of the RNAPs is linearly related to the density, and the velocity of those affected RNAPs increases with decreasing density. Therefore, the trailing RNAPs that are stopped by the first pause but escape effects of the second 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 because of the lower density directly in front of them as they travel through the rarefaction wave that occurs.

The main conclusion that we draw from the results of Figs. 5.1a-5.1b is that if a pause occurs at a particular location, then the *background density* or the overall crowding or congestion of RNAPs on the DNA strand can have a significant effect on how often that pause should occur in order to minimize the average amount of delay experienced by the individual polymerases. That is, the minimum delay value occurs at a different choice of the parameter value determining the time between pauses for each case of z_0. In addition, for each choice of z_0, once the time between the pauses exceeds a critical threshold, the average delay approaches an “equilibrium value,” and the delay is insensitive to very small changes in the time between the two pauses once the parameter is sampled from that region. For example, when one chooses a background density of $z_0 = 0.3$, if the nominal values of the parameter i determining the time between the two pauses is selected anywhere in the region $(2, 6)$, the average delay value will be insensitive to small changes in that parameter choice. However, Fig. 5.1b indicates that for the case of higher background density of $z_0 = 0.4$, when the nominal parameter value of i is located in the interval $(2, 4)$, the value of time delay will be sensitive to small changes in the parameter controlling the time between the pauses. Indeed, the more congested case would require the parameter to be chosen to be larger than 5 in order to guarantee that the delay value is insensitive to small changes in the time between pauses. From the biological perspective, increases in the average delay per RNAP are directly related to decreases in the overall transcriptional output of the RNAPs, thereby decreasing the mRNA production of the gene.

5.2. Parameterized location of a second pause

In the previous section, the spatial location of the two pauses is fixed and the time between them is varied. We now show 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 $0.6 < t < 0.65$. 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 the time interval is chosen so that for each choice of the parameter value, every polymerase affected by the pauses has reached termination and the density is back to the reference case density value, $z_0 = 0.31$ for this case. A local maximum and two local minimums in the delay are observed in

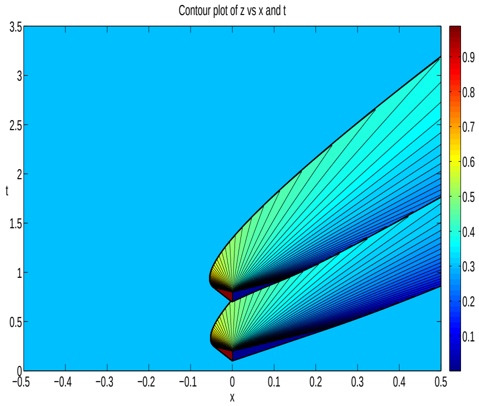
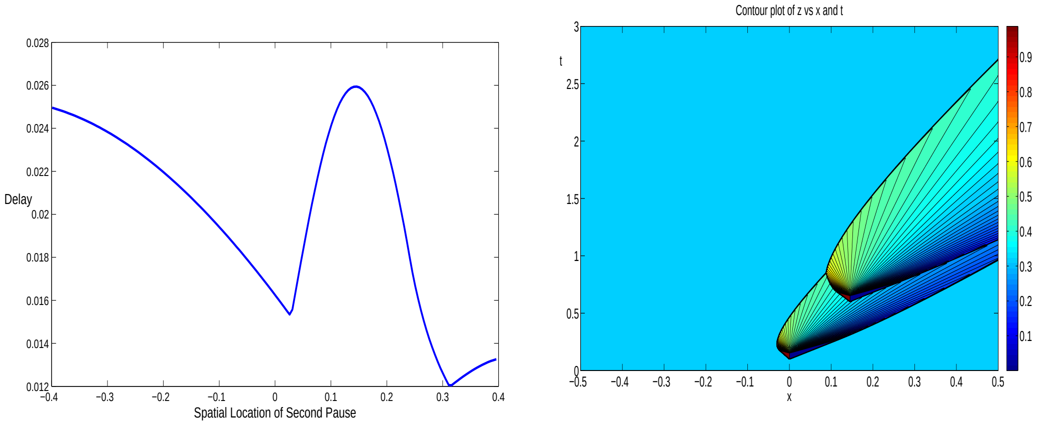


Fig. 5.2.: Contour plot of the PDE simulation corresponding to the minimum delay value in Fig. 5.1b with an initial density value of $z_0 = 0.3$.

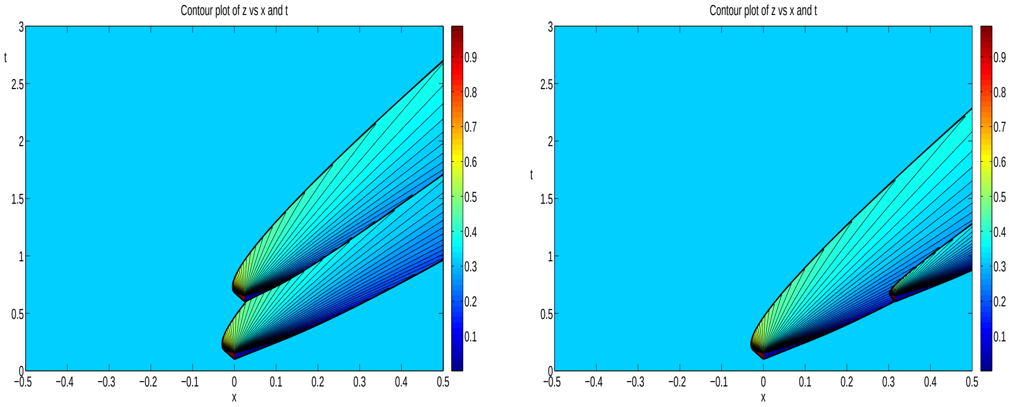


(a) Average delay as a function of the spatial location of the second pause.

(b) 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. 5.3.: Results of parameter study in Section 5.2.

Fig. 5.3a. 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 in a given time period. Figs. 5.4a and 5.4b include the contour plots of the simulations corresponding to the local minimum delay values indicating 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. Fig. 5.3b includes the contour plot of the model simulation corresponding to the maximum delay value indicating that it occurs when the characteristic created by the release of the second pause intersects the shock emanating from the first pause. We might also note that the actual percentage of the average delay only ranges from less than 1% to approximately 2% of the average transcription time for the case that $z_0 = 0.31$. As shown in Figs. 5.1a and 5.1b, the initial density as well as pause location parameterized in time and space can influence the average delay, and as the DNA strand becomes more crowded, then we would expect the location of the second pause to affect the average delay number



(a) Density and location at the first local minimum.

(b) Density and location at the second local minimum.

Fig. 5.4.: Contour plot of the density $z(x,t)$ from the parameter study in Section 5.2 where the location of the second pause corresponds to a local minimum value of the delay.

more significantly. However, it is unclear which has a larger effect on the average time delay experienced per polymerase. Based on results here, the low density regimes may show the same amount of sensitivity to small changes in location and time durations of pauses; however, the results in Fig. 5.1b lead us to hypothesize that for the case where the DNA strand is more crowded (that is, $0.3 < z_0 < 0.5$), the average delay per polymerase may be more sensitive to small changes in duration times of the pauses. Investigations where parameterizations varying both spatial locations and time durations are simultaneously considered are currently underway.

6. Conclusions and Future Work

This work presents a careful numerical study of a nonlinear PDE model that describes DNA transcription in the presence of short pauses that are located throughout the spatial and time domains. In highly transcribed genes such as the *rrn* gene in E.coli under high growth conditions, the simulation of the interaction of high density of polymerases with frequent pauses requires employment of 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 this 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. Here, it is assumed that the overall production of ribosomes is directly related to the measured amount of RNAPs that transcribe the DNA strand over a given amount of time. Convergence of order $\mathcal{O}(\Delta x)$ for the numerical scheme is observed for both the density variable and the subsequent delay computation in those simple cases where we are able to test against an analytical solution.

A rigorous study of the computational results of the PDE model with many pauses randomly distributed throughout the domain is difficult and will require a statistical approach. This type of study is the subject of ongoing research. We make the first steps toward a different approach by considering a thorough parameter sensitivity study for simple models such as the two pause model described in Sections 5.1 and 5.2. Important parameters include the spatial location and 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 functions of cells.

7. Appendix

In this section, we begin by discussing the simple model problems that allow us to incorporate pause locations into PDE model. Consider the conservation law given by

$$z_t + f_x = 0,$$

where $z(x,t)$ is the traffic density and f is the flux function. Assume that velocity is explicitly dependent on density only so that the flow is

$$f = f(z) = \beta_m \left(1 - \frac{z}{z_m}\right) z,$$

where β_m and z_m are parameters indicating maximum values of the velocity and the density, respectively.

7.1. Elongation after a long pause

First we consider a mathematical model which describes the behavior of the system once the bottleneck nucleotide releases the lead polymerase that is paused at one site along the DNA strand. For this discussion, the spatial location of the pause is fixed at $x = 0$.

$$z_t + \left(\beta_m \left(1 - \frac{z}{z_m}\right) z\right)_x = 0 \quad x \in (-L/2, L/2), \quad t > 0, \tag{7.1a}$$

$$z(x,0) = \begin{cases} z_m & x < 0, \\ 0 & x > 0, \end{cases} \tag{7.1b}$$

$$z(-L/2,t) = z_m. \tag{7.1c}$$

The initial condition in Eq. (7.1b) assumes that the duration of the pause is sufficiently long so that the RNAPs back up along the entire length of the DNA strand behind the paused polymerase and that the density is at its maximum value, z_m, behind the bottleneck. The boundary condition in (7.1c) assumes that the initiation rate is large enough that the density remains at its maximum value once the slow nucleotide releases the paused ribosome and allows elongation to continue.

Using the method of characteristics, one can develop a closed-form solution to (7.1). The characteristics for the PDE are calculated by solving

$$\frac{dz}{dt} = 0,$$
$$\frac{dx}{dt} = \frac{df}{dz} = f'(z).$$

Since z is constant along these curves, the characteristics $x(t)$ are straight lines with slope $f'(z)$ where z is the value of the initial condition. If the characteristic passing through (x, t) interests the x-axis at $x_0 < 0$, the solution is $z = z_m$ and the characteristic has slope $f'(z_m) = -\beta_m$. The equation of this characteristic is

$$x_{cl} = -\beta_m t + x_0. \tag{7.2}$$

If the characteristic through (x, t) intersects the x-axis at $x_0 > 0$, the value of the solution at that point is $z = 0$ and the corresponding characteristic has slope $f'(0) = \beta_m$. The equation of the characteristic is

$$x_{cr} = \beta_m t + x_0. \tag{7.3}$$

To examine the area of the xt-plane that is not covered by the characteristic curves calculated previously, consider

$$\frac{dx}{dt} = \frac{df}{dz}, \quad x(0) = 0, \qquad \text{for } -\beta_m t < x < \beta_m t.$$

Since $\frac{dz}{dt} = 0$, z is constant along characteristics and therefore $\frac{df}{dz}$ is also a constant. Therefore

$$x = \frac{df}{dz} t, \quad \Longrightarrow \quad \frac{x}{t} = \frac{df}{dz} = \beta_m \left(1 - \frac{2z}{z_m}\right).$$

Solving this equation for z gives

$$z = \frac{z_m}{2} - \frac{z_m x}{2\beta_m t}. \tag{7.4}$$

Then the expression for the solution to (7.1a)-(7.1c) is

$$z(x,t) = \begin{cases} z_m & x < -\beta_m t, \\ \dfrac{z_m}{2} - \dfrac{z_m x}{2\beta_m t} & -\beta_m t < x < \beta_m t, \\ 0 & \beta_m t < x. \end{cases}$$

This provides us with an expression for the solution to the nonlinear conservation law beginning at the instant that the slow nucleotide releases its paused RNAPs and elongation is allowed to continue. In the next subsection, we build on this work and use the previous expression to construct a closed-form solution that incorporates the behavior of density during the time period of the pause.

7.2. Incorporation of the pause

In this section, we examine the PDE in (7.1a) where the constant maximum velocity, β_m, is replaced by piecewise constant function

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

Consider

$$z_t + \left(\beta(x,t)\left(1-\frac{z}{z_m}\right)z\right)_x = 0, \quad x \in (-L/2, L/2), \quad t > 0, \tag{7.5a}$$

$$z(x,0) = z_0, \tag{7.5b}$$

$$z(-L/2,t) = z_0, \tag{7.5c}$$

where $\beta(x,t)$ represents a velocity function which assumes a maximum value of β_m 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, the velocity function is set to 0 and no elongation is allowed to take place at this spatial location. We use this formulation to model the location of a bottleneck codon (or a traffic light) at $x = 0$, beginning at time $t = 0$ and ending at time $t = \tau$. Assume that $z_0 < \frac{1}{2}$ so that the initial density is low to moderate; correspondingly, the flow is below its maximum value prior to an individual polymerase encountering a line of paused RNAPs.

For $0 < t < \tau$, there are three shocks emanating from $x = 0$. To the left of the slow nucleotide, the density jumps from z_0 to z_m as the RNAPs transcribe the DNA strand until they are stopped behind the paused RNAP. Therefore there is a shock, denoted x_1, with speed and initial condition satisfying

$$\frac{dx_1}{dt} = \frac{-\beta_m\left(1-\frac{z_0}{z_m}\right)z_0}{z_m - z_0} = \frac{-\beta_m z_0}{z_m}, \quad x_1(0) = 0,$$

resulting in

$$x_1 = \frac{-\beta_m z_0}{z_m}t. \tag{7.6}$$

At x=0, the density jumps from z_m to 0 and therefore creates a shock, x_2, with speed determined by the IVP

$$\frac{dx_2}{dt} = 0, \quad x_2(0) = 0,$$

resulting in the shock

$$x_2 = 0.$$

To the right of the pause, the density jumps from 0 to z_0 as the RNAPs in front of the pause continue the elongation process, and this also creates a shock, x_3, with speed determined by the IVP

$$\begin{aligned}\frac{dx_3}{dt} &= \frac{-\beta_m\left(1-\frac{z_0}{z_m}\right)z_0}{-z_0}\\ &= \frac{\beta_m(z_m - z_0)}{z_m}, \quad x_3(0) = 0,\end{aligned}$$

and resulting in the expression

$$x_3 = \frac{\beta_m(z_m - z_0)}{z_m}t. \tag{7.7}$$

Using these shocks, a closed-form representation of the solution to the PDE in (7.5a)-(7.5c) for $0 \leq t < \tau$ is given by

$$z(x,t) = \begin{cases} z_0 & x < \dfrac{-\beta_m z_0}{z_m} t, \\ z_m & \dfrac{-\beta_m z_0}{z_m} t < x < 0, \\ 0 & 0 < x < \dfrac{\beta_m (z_m - z_0)}{z_m} t, \\ z_0 & \dfrac{\beta_m (z_m - z_0)}{z_m} t < x. \end{cases}$$

The above expression for $z(x,t)$ describes the behavior of polymerase density during the time that the pause occurs. Next we also describe the solution for $t \geq \tau$ by making use of work from the previous section. At $t = \tau$, the pause at the slow nucleotide releases the lead polymerase resulting in a problem similar to the situation considered in Section 7.1. As the pause releases and the RNAPs begin to elongate at $x = 0$, a rarefaction wave is created from the maximum density, $z = z_m$, to the minimum density, $z = 0$. Using Eq. (7.4) of Section 7.1, the density in this region is

$$z_w = \frac{z_m}{2} - \frac{z_m x}{2\beta_m (t - \tau)}. \tag{7.8}$$

Here the density jumps from z_0 to z_w; therefore, there is a shock, x_s, with speed determined by the equation

$$\frac{dx_s}{dt} = \frac{f(z_w) - f(z_0)}{z_w - z_0} = 1 - (z_w + z_0).$$

Hence,

$$\frac{dx_s}{dt} = \beta_m \left[\frac{1}{2} + \frac{x_s}{2\beta_m (t - \tau)} - \frac{z_0}{z_m} \right],$$

which is a first order linear ODE.

An integrating factor is $\mu = (t - \tau)^{-1/2}$, and the corresponding general solution to the ODE is

$$x_s = 2\beta_m \left[\frac{1}{2} - \frac{z_0}{z_m} \right] (t - \tau) + c(t - \tau)^{1/2}, \tag{7.9}$$

where c is an arbitrary constant of integration that is resolved by an initial condition.

To find the appropriate initial condition for the left shock, find the time value t_1 such that the characteristic (7.2) to the left of the rarefaction wave (shifted in time by τ) intersects the shock (7.6) emanating to the left of the pause. That is, we find t_1 such that

$$-\beta_m (t_1 - \tau) = \frac{-z_0 \beta_m t_1}{z_m},$$

then

$$t_1 = \frac{z_m \tau}{z_m - z_0}.$$

To find the left shock, solve the IVP

$$\frac{dx_{sl}}{dt} = \beta_m \left[\frac{1}{2} + \frac{x_{sl}}{2\beta_m (t - \tau)} - \frac{z_0}{z_m} \right],$$

$$x_{sl}(t_1) = \frac{-z_0 \beta_m t_1}{z_m} = \frac{-z_0 \beta_m \tau}{z_m - z_0}.$$

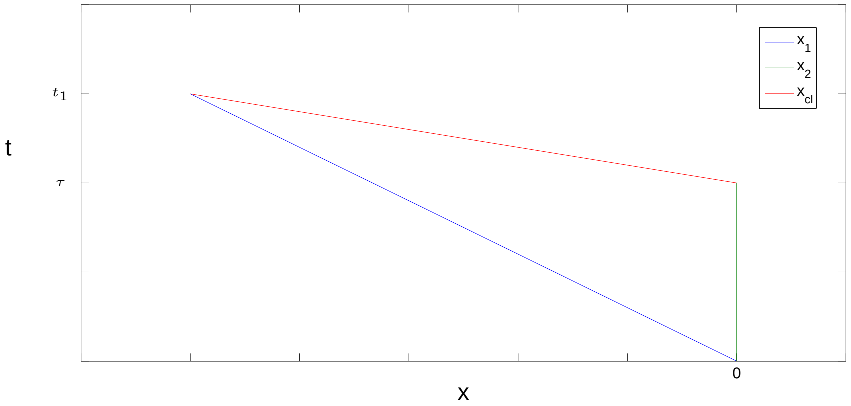


Fig. 7.1.: Plot of two shocks and a characteristic. The blue and green curves are shocks emanating from (0, 0). The red curve is a characteristic bordering the rarefaction wave, and the intersection of the characteristic and shock creates the shock, x_{sl}, at the time, t_1.

Imposing the initial condition yields

$$c = -2\beta_m \left[\frac{z_0\tau}{z_m - z_0}\right]^{1/2} \left(1 - \frac{z_0}{z_m}\right),$$

and the equation for the left shock is given by

$$x_{sl} = \beta_m \left[1 - \frac{2z_0}{z_m}\right](t-\tau) - 2\beta_m \left[\frac{z_0\tau(t-\tau)}{z_m - z_0}\right]^{1/2} \left(1 - \frac{z_0}{z_m}\right), \quad t \geq t_1.$$

Similarly, to find the appropriate initial condition for the right shock, let t_2 be the time that the characteristic (7.3) to the right of the rarefaction wave (shifted in time by τ) intersects the shock (7.7) emanating to the right of the pause. Find t_2 such that

$$\beta_m(t_2 - \tau) = \frac{\beta_m(z_m - z_0)}{z_m} t_2,$$

then

$$t_2 = \frac{z_m\tau}{z_0}.$$

To find the right shock, one solves the IVP

$$\frac{dx_{sr}}{dt} = \beta_m \left[\frac{1}{2} + \frac{x_{sr}}{2\beta_m(t-\tau)} - \frac{z_0}{z_m}\right], \quad x_{sr}(t_2) = \beta_m(t_2 - \tau) = \beta_m\left(\frac{z_m}{z_0} - 1\right)\tau.$$

Imposing the initial condition yields

$$c = \frac{2z_0\beta_m}{z_m}\left(\frac{z_m}{z_0} - 1\right)^{1/2} \tau^{1/2},$$

and the equation for the right shock is given by

$$x_{sr} = 2\beta_m\left[\frac{1}{2} - \frac{z_0}{z_m}\right](t-\tau) + \frac{2z_0\beta_m}{z_m}\left[\left(\frac{z_m}{z_0} - 1\right)\tau(t-\tau)\right]^{1/2}, \quad t \geq t_2.$$

Fig. 7.3 includes a plot of the five computed shocks and the two characteristics that border the rarefaction wave. The plot also labels each region with the solution value for reference when building the piecewise defined true solution.

7.3. Pause at a parameterized interval of time

One can also solve the related problem, where the pause at the bottleneck codon site does not occur at the initial time 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

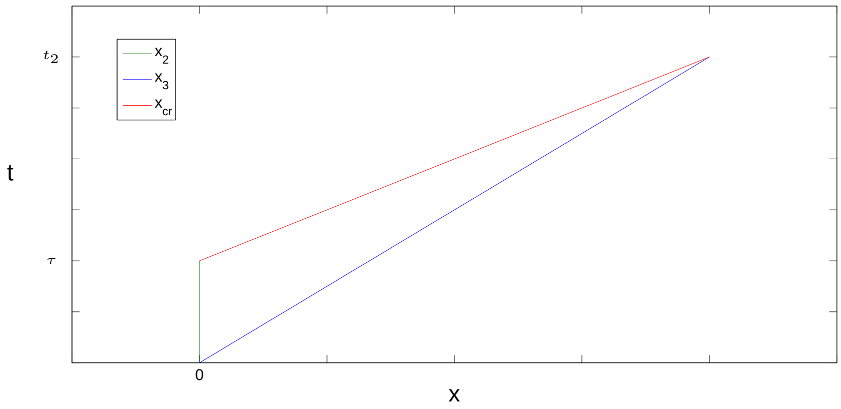


Fig. 7.2.: Plot of two shocks and a characteristic. The blue and green curves are shocks emanating from (0, 0). The red curve is a characteristic bordering the rarefaction wave. The intersection of the characteristic and shock creates the shock, x_{sr}, at the time, t_2.

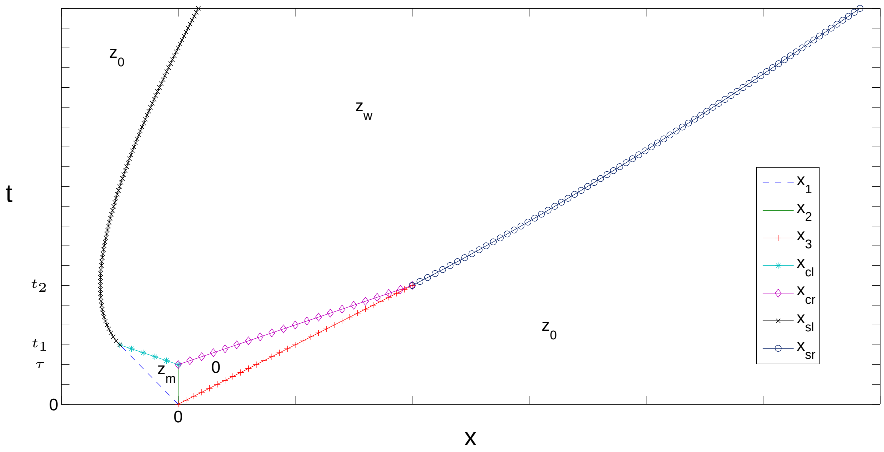


Fig. 7.3.: Plot of all shocks and bordering characteristics. The values of the solution are labeled in each region including the constant values 0, z_0 and z_m and the rarefaction wave defined by z_w (7.8).

function β. Hence we solve the PDE in (7.5a) - (7.5c) but define $\beta(x,t)$ as

$$\beta(x,t)=\begin{cases} 0 & x=0; \quad \zeta<t<\xi \\ \beta_m & \text{otherwise.} \end{cases}$$

The solution can be found by merely shifting the solution in the previous subsection in time by ζ. With $z_m=1$, $\beta_m=1$,

$$dt=\xi-\zeta, \quad t_1=\frac{dt}{1-z_0}, \quad t_2=\frac{dt}{z_0}$$

and $z_0=0.31$, the closed-form solution to (7.5a)-(7.5c) is given by

$$z(x,t)=\begin{cases} 0.31 & t<\zeta, \\ 0.31 & x<-0.31(t-\zeta); \quad t<t_1+\zeta, \\ 1 & -0.31(t-\zeta)<x<0; \quad t<\xi, \\ 1 & -0.31(t-\zeta)<x<-(t-\xi); \quad \xi<t<t_1+\zeta, \\ \dfrac{1}{2}-\dfrac{x}{2(t-\xi)} & -(t-\xi)<x<0; \quad \xi<t<t_1+\zeta, \\ 0.31 & x<0.38(t-\xi)-2\left[\frac{3.1(t-\xi)}{69}\right]^{1/2}(0.69); \quad t\geq t_1+\zeta, \\ \dfrac{1}{2}-\dfrac{x}{2(t-\xi)} & 0.38(t-\xi)-2\left[\frac{3.1(t-\xi)}{69}\right]^{1/2}(0.69)<x<0; \quad t\geq t_1+\zeta, \\ 0 & 0<x<(0.69)(t-\zeta); \quad t<\xi, \\ \dfrac{1}{2}-\dfrac{x}{2(t-\xi)} & 0<x<t-\xi; \quad \xi<t<t_2+\zeta, \\ 0 & t-\xi<x<(0.69)(t-\zeta); \quad \xi<t<t_2+\zeta, \\ 0.31 & (0.69)(t-\zeta)<x; \quad t<t_2+\zeta, \\ \dfrac{1}{2}-\dfrac{x}{2(t-\xi)} & 0<x<0.38(t-\xi)+0.62\left[\left(\frac{69}{31}\right)(t-\xi)(0.1)\right]^{1/2}; \quad t\geq t_2+\zeta, \\ 0.31 & 0.38(t-\xi)+0.62\left[\left(\frac{69}{31}\right)(t-\xi)(0.1)\right]^{1/2}<x; \quad t\geq t_2+\zeta. \end{cases} \tag{7.10}$$

Acknowledgments. The authors wish to acknowledge support from the NSF under grant DMS-1226213.

References

[1] H. Bremer and P. P. Dennis, *Escherichia coli and Salmonella typhi-murium: Cellular and Molecular Biology,* Chapter - Modulation of chemical composition and other parameters of the cell by growth rate, eds. Neidhardt EA (Sinauer, Sunderland, MA) 2nd Ed., College Station, Texas, (1996), 421.

[2] B. Cockburn, Discontinuous Galerkin methods for convection-dominated problems, In *High-order methods for computational physics,* **Volume 9** of Lect. Notes Comp. Sci. Eng. Springer, Berlin, 1999, 69-224.

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

[4] Bernardo Cockburn and Chi-Wang Shu, TVB Runge-Kutta Local projection discontinuous Galerkin finite element method for conservation laws II: General Framework, *Math. Comp.,* **52**:186 (1989), 411-435.

[5] C. Condon, S. French, C. Squires, and C. L. Squires, Depletion of functional ribosomal RNA operons in Escherichia coli causes increased expression of the remaining intact copies, *The EMBO Journal,* **12**:11 (1993), 4305-4315.

[6] L. Davis, T. Gedeon, J. Gedeon, and J. Thorenson, A traffic flow model for bio-polymerization processes, *J. Math. Bio.,* **68**:3 (2014), 667-700.

[7] P.P. Dennis, M. Ehrenberg, D. Fange, and H. Bremer, Varying rate of RNA chain elongation during rrn transcription in escherichia coli. *J. Bacter.,* **191**:11 (2009), 3740-3746.

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

[9] R. Haberman, Mathematical models: Mechanical vibrations, population dynamics, and traffic flow, volume 21 of *Classics in Applied Mathematics,* SIAM, Philadelphia, PA, 1998 (Reprint of the 1977 original).

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

[11] M.L. Kireeva and M. Kashlev, Mechanism of sequence-specific pausing of bacterial RNA polymerase, *PNAS,* **106**:22 (2009), 8900-8905.

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

[13] R. Landick, Transcriptional pausing without backtracking, *PNAS,* **106**:22 (2009), 8797-8798.

[14] Randall LeVeque, Finite Volume Methods for Hyperbolic Problems, Cambridge University Press, New York, NY, 2002.

[15] M.J. Lighthill and G.B. 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 (1955), 317-345.

[16] R.-X. Liu, H. Li, and Z.-F. Wang, The discontinuous finite element method for red-and-green light models for the traffic flow. *Math. Comput. Simulation,* **56**:1 (2001), 55-67.

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

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

[19] P. Zhang and R.-X. Liu, Generalization of Runge-Kutta discontinuous Galerkin method to LWR traffic flow model with inhomogeneous road conditions, *Numer. Methods Partial Differential Equations,* **21**:1 (2005), 80-88.

[20] R.K.P. Zia, J.J. Dong and B. Schmittmann, Modeling translation in protein synthesis with TASEP: a tutorial and recent developments, *J. Stat. Phys.,* **144**:2 (2011), 405-428.