Controllability Analysis and Control Synthesis for the Ribosome Flow Model

Yoram Zarai, Michael Margaliot, Eduardo D. Sontag, and Tamir Tuller

Abstract—The ribosomal density along different parts of the coding regions of the mRNA molecule affects various fundamental intracellular phenomena including: protein production rates, global ribosome allocation and organismal fitness, ribosomal drop off, co-translational protein folding, mRNA degradation, and more. Thus, regulating translation in order to obtain a *desired* ribosomal profile along the mRNA molecule is an important biological problem. We study this problem by using a dynamical model for mRNA translation, called the ribosome flow model (RFM). In the RFM, the mRNA molecule is modeled as an ordered chain of n sites. The RFM includes n state-variables describing the ribosomal density profile along the mRNA molecule, and the transition rates from each site to the next are controlled by $n+1$ positive constants. To study the problem of controlling the density profile, we consider some or all of the transition rates as time-varying controls. We consider the following problem: given an initial and a desired ribosomal density profile in the RFM, determine the time-varying values of the transition rates that steer the system to the desired density profile, if they exist. More specifically, we consider two control problems. In the first, all transition rates can be regulated separately, and the goal is to steer the ribosomal density profile and the protein production rate from a given initial value to a desired value. In the second problem, one or more transition rates are jointly regulated by a single scalar control, and the goal is to steer the production rate to a desired value within a certain set of feasible values. In the first case, we show that the system is controllable, i.e., the control is powerful enough to steer the system to any desired value in finite time, and provide simple closed-form expressions for *constant* positive control functions (or transition rates) that asymptotically steer the system to the desired value. In the second case, we show that the system is controllable, and provide a simple algorithm for determining the *constant* positive control value that asymptotically steers the system to the desired value. We discuss some of the biological implications of these results.

Index Terms—Systems biology, synthetic biology, gene translation, ribosomal density profile, controllability, asymptotic controllability, accessibility, control-affine systems, Lie-algebra, control synthesis, ribosome flow model

1 Introduction

The process in which the genetic information coded in the DNA is transformed into functional proteins is called *gene expression*. It consists of two major steps: *transcription* of the DNA code into messenger RNA (mRNA) by RNA polymerase, and *translation* of the mRNA into proteins. During the translation step, complex macro-molecules called ribosomes unidirectionally traverse the mRNA, decoding it codon by codon into a corresponding chain of amino-acids that is folded co-translationally and post-translationally to become a functional protein. The rate in which proteins are produced during the translation step is called the *protein translation rate* or *protein production rate*.

Translation takes place in all living organisms and all tissues under almost all conditions. Thus, developing a better understanding of how translation is regulated has important implications to many scientific disciplines, including medicine, evolutionary biology, and synthetic biology. Developing and analyzing computational models of translation may provide important insights on this biological process. Such models can also aid in integrating and analyzing the rapidly increasing experimental findings related to translation (see, e.g., [7], [9], [12], [46], [54], [62], [64]).

Controlling the expression of heterologous genes in a host organism in order to synthesize new proteins, or to improve certain aspects of the host fitness, is an essential challenge in biotechnology and synthetic biology [3], [4], [37], [50], [63]. Computational models of translation are particularly important in this context, as they allow simulating and analyzing the effect of various manipulations of the gene expression machinery and/or the genetic material, and can thus save considerable time and effort by guiding biologists towards promising experimental directions.

The ribosome flow along the mRNA is regulated by various translation factors (e.g., initiation and elongation factors, tRNA and Aminoacyl tRNA synthetase concentrations, and amino-acid concentrations) in order to achieve both a suitable ribosomal density profile along the mRNA, and a desired protein production rate. Indeed, it is known that the ribosomal density profile and the induced ribosome speed profile along the mRNA molecule can affect various

- *Y. Zarai is with the Department of Biomedical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. E-mail: yoramzar@mail.tau.ac.il.*
- *M. Margaliot is with the School of Electrical Engineering and the Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv 69978, Israel. E-mail: michaelm@eng.tau.ac.il.*
- *E.D. Sontag is with the Department of Mathematics and Cancer Center of New Jersey, Rutgers University, Piscataway, NJ 08854. E-mail: eduardo.sontag@gmail.com.*
- *T. Tuller is with the Department of Biomedical Engineering and the Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv 69978, Israel. E-mail: tamirtul@gmail.com.*

Manuscript received 7 Feb. 2016; revised 28 Jan. 2017; accepted 17 May 2017. Date of publication 23 May 2017; date of current version 6 Aug. 2018. (Corresponding author: Tamir Tuller.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2017.2707420

fundamental intracellular phenomena. For example, it is known that the folding of translated proteins may take place co-translationally, and inaccurate translation speed can contribute to protein mis-folding [15], [28], [70]. The ribosome density profile also affects the degradation of mRNA, ribosomal collisions, abortion and allocation, transcription, and more [15], [17], [28], [29], [45], [63], [70].

Thus, a natural question is whether it is possible, by controlling the transition rates along the mRNA, to steer the ribosome density along the mRNA molecule from any initial profile to any desired profile in finite time, and if so, how. In the language of control theory, the question is whether the system is *controllable* (see, e.g., [58]), and if so, how to solve the control synthesis problem. We note that controllability of networked systems is recently attracting considerable interest (see e.g., [31]). Controllability of such networks depends on the interplay between two factors: (1) the network's topology, and (2) the dynamical rules describing the behavior at each network node. When studying real-world networks, many of the parameter values in the network are not known explicitly. The network is said to be *structurally stable* if it will be controllable for almost every random selection of parameter values [30], [36], [56], [58]. An important problem in this context is to determine a minimal set of "driver nodes" within the network such that controlling these nodes makes the entire network controllable or structurally controllable (see e.g., [31], [39]).

Controllability of mRNA translation is also important in synthetic biology, e.g., in order to design cis or trans intracellular elements that yield a desired ribosome density profile (or to determine if such a design is possible). Another related question arises in evolutionary systems biology, namely, determine if a certain translation-related phenotype can be obtained by evolution.

The ribosome density profile is also related to cancer evolution. Indeed, it is well-known that cancerous cells undergo evolution that modulates their translation regime. It has been suggested that various mutations that accumulate during tumorigenesis may affect both translation initiation [21], [32] and elongation [60], [65] of genes related to cell proliferation, metabolism, and invasion. Specifically, the results reported in [21] support the conjecture that cancerous mutations can significantly change the ribosome density profile on the mRNAs of dozens of genes.

The standard mathematical model for ribosome flow is the *totally asymmetric simple exclusion process* (TASEP) [55], [71]. In this model, particles hop unidirectionally along an ordered lattice of L sites. Every site can be either free or occupied by a particle, and a particle can only hop to a free site. This simple exclusion principle models particles that have "volume" and thus cannot overtake one other. The hops are stochastic and the rate of hoping from site i to site $i+1$ is denoted by γ_i. A particle can hop to [from] the first [last] site of the lattice at a rate α [β]. The flow through the lattice converges to a steady-state value that depends on L and the parameters $\alpha, \gamma_1, \ldots, \gamma_{L-1}, \beta$. In the context of translation, the lattice models the mRNA molecule, the particles are ribosomes, and simple exclusion means that a ribosome cannot overtake a ribosome in front of it. TASEP has become a fundamental model in non-equilibrium statistical mechanics, and has been applied to model numerous natural and artificial processes [53].

The *ribosome flow model* (RFM) [49] is a *deterministic* model for mRNA translation that can be derived via a dynamic mean-field approximation of TASEP [53, section 4.9.7] [5, p. R345]. In the RFM, mRNA molecules are coarse-grained into n consecutive sites of codons (or groups of codons). The state variable $x_i(t) : \mathbb{R}_+ \to [0,1]$, $i = 1, \ldots, n$, describes the normalized ribosomal occupancy level (or density) of site i at time t, where $x_i(t) = 1$ [$x_i(t) = 0$] indicates that site i is completely full [empty] at time t. Thus, the vector $[\, x_1(t) \ldots x_n(t) \,]'$ describes the complete *ribosomal density profile* along the mRNA molecule at time t. A variable denoted $R(t)$ describes the *protein production rate* at time t. A non-negative parameter λ_i, $i = 0, \ldots, n$, controls the transition rate from site i to site $i+1$, where λ_0 [λ_n] is the initiation [exit] rate.

In order to better understand how translation is regulated, we consider the RFM with some or all of the constant transition rates replaced by time-varying control functions that take non-negative values for all time t. The idea here is that we can manipulate these functions as desired.

We consider two control problems. In the first, all the $n+1$ λ_is are replaced by control functions and the problem is to manipulate these functions such that both the ribosomal density profile and the production rate are steered from a given initial value to a desired value. We use the term "augmented profile" to indicate the combination of the ribosomal density profile and the production rate.

In the second control problem, we assume that all the rates belonging to some *subset* of the rates are *jointly* replaced by a single, scalar control $u(t)$. We define a set of "relevant" possible production rates and the problem is to determine $u(t)$ such that the production rate is steered to a desired value in this set. Note that in the first problem the $(n+1)$-dimensional vector describing the augmented profile is controlled using $n+1$ control functions, and in the second problem one variable is controlled using a scalar control.

We show that in both cases the resulting control system is *controllable*, i.e., the control is always "powerful" enough to steer the system from any initial state to any desired state in some finite time T. We also show that there always exists a control that steers the system as desired, and is the time concatenation of two controls

$$u(t) = \begin{cases} v, & t \in [0, T-\varepsilon), \\ w(t), & t \in [T-\varepsilon, T], \end{cases} \tag{1}$$

with $\varepsilon > 0$ and very small. The constant control v is given in a *simple and explicit* expression that depends only on the desired final state. It guarantees that this state becomes the unique attracting steady-state ribosomal density and production rate of the RFM dynamics. For example, in the problem of controlling the density profile and the production rate to desired final values x^f and R^f, respectively ("f" for final), the solution of the controlled RFM for *any* initial condition $x(0)$ and $R(0)$ satisfies

$$\begin{aligned} \lim_{t\to\infty} x(t,v) &= x^f, \\ \lim_{t\to\infty} R(t,v) &= R^f. \end{aligned} \tag{2}$$

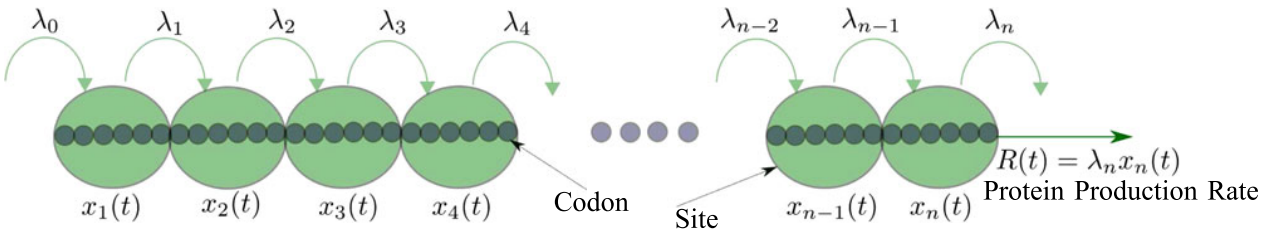


Fig. 1. The RFM models a chain of n sites of codons (or groups of codons). The state variable $x_i(t) \in [0,1]$ represents the normalized ribosome occupancy at site i at time t. The elongation rate from site i to site $i+1$ is λ_i, with λ_0 [λ_n] denoting the initiation [exit] rate. The production rate at time t is $R(t) = \lambda_n x_n(t)$.

This means that for all practical reasons, one may simply apply the constant control $u(t) \equiv v$ for all $t \geq 0$. Note that (2) means that the exact values of $x(0)$ and $R(0)$, i.e., the initial values of the density profile and production rate, are actually not needed. This is important, as accurately measuring $x(0)$ and $R(0)$ in practice may be difficult. The control $w(t)$ in (1) is needed only to guarantee that $x(T) = x^f$ and $R(T) = R^f$ at the *finite* time T. The existence of such a $w(t)$ follows from Lie-algebraic accessibility arguments, but $w(t)$ is not given explicitly.

Different aspects of translation regulation, usually under natural conditions, have been studied before (see, for example, [22]). There are also several studies on experimental and computational heuristics for mRNA translation engineering and optimization (see, for example, [52], [59]), and studies related to the way translation regulation is encoded in the transcript (e.g., [42], [72]). However, to the best of our knowledge, this is the first study on controllability and control synthesis in a realistic dynamical model for translation. Also, previous studies on translation optimization only considered protein levels or production rate (e.g., [52]), but not the problem of controlling the entire profile of ribosome densities via changing the codon decoding rates, as is done here.

The remainder of this paper is organized as follows. The following section provides a brief overview of the RFM and its generalizations into a control system. In order to make this paper accessible to a larger audience, Appendix A provides a very brief review of controllability, while demonstrating some of the concepts using the RFM. Section 3 presents our main results on the controlled RFM. We also discuss the biological ramifications of our results. To streamline the presentation, all the proofs are placed in Appendix B. We use standard notation. Vectors [matrices] are denoted by small [capital] letters. For a vector $x \in \mathbb{R}^n$, x_i is the ith entry of x, and x' is the transpose of x. $\mathbb{R}^n_+$ [$\mathbb{R}^n_{++}$] is the set all n-tuples of nonnegative [strictly positive] real numbers.

2 Ribosome Flow Model

In this section, we quickly review the RFM and describe its generalizations into a control system. The dynamics of the RFM with n sites is given by n nonlinear first-order ordinary differential equations

$$\begin{aligned}
\dot{x}_1 &= \lambda_0(1-x_1) - \lambda_1 x_1(1-x_2),\\
\dot{x}_2 &= \lambda_1 x_1(1-x_2) - \lambda_2 x_2(1-x_3),\\
\dot{x}_3 &= \lambda_2 x_2(1-x_3) - \lambda_3 x_3(1-x_4),\\
&\vdots\\
\dot{x}_{n-1} &= \lambda_{n-2} x_{n-2}(1-x_{n-1}) - \lambda_{n-1} x_{n-1}(1-x_n),\\
\dot{x}_n &= \lambda_{n-1} x_{n-1}(1-x_n) - \lambda_n x_n.
\end{aligned} \tag{3}$$

If we define $x_0(t) := 1$ and $x_{n+1}(t) := 0$ then (3) can be written more succinctly as

$$\dot{x}_i = \lambda_{i-1} x_{i-1}(1-x_i) - \lambda_i x_i(1-x_{i+1}), \quad i = 1, \ldots, n. \tag{4}$$

Recall that the state variable $x_i(t) : \mathbb{R}_+ \to [0,1]$ describes the normalized ribosomal occupancy level (or density) at site i at time t, where $x_i(t) = 1$ [$x_i(t) = 0$] indicates that site i is completely full [empty] at time t. Eq. (4) can be explained as follows. The flow of ribosomes from site i to site $i+1$ is $\lambda_i x_i(t)(1 - x_{i+1}(t))$. This flow is proportional to $x_i(t)$, i.e., it increases with the occupancy level at site i, and to $(1 - x_{i+1}(t))$, i.e., it decreases as site $i+1$ becomes fuller. This corresponds to a "soft" version of the simple exclusion principle in TASEP. Note that the maximal possible flow from site i to site $i+1$ is the transition rate λ_i. Eq. (4) thus states that the time derivative of state-variable x_i is the flow entering site i from site $i-1$, minus the flow exiting site i to site $i+1$.

The ribosome exit rate from site n at time t is equal to the protein production rate at time t, and is denoted by $R(t) := \lambda_n x_n(t)$ (see Fig. 1). Note that x_i is dimensionless, and every rate λ_i has units of $1/\text{time}$.

A system where each state variable describes the amount of "material" in some compartment, and the dynamics describes the flow of material between the compartments and also to/from the surrounding environment is called a *compartmental system* [24]. Compartmental systems proved to be useful models in various biological domains including physiology, pharmacokinetics, population dynamics, and epidemiology [6], [20], [23]. The RFM is thus a nonlinear compartmental model, with x_i denoting the normalized amount of "material" in compartment i, and the flow follows a "soft" simple exclusion principle. The controllability of *linear* compartmental systems has been addressed in several papers [19], [25].

Let $x(t,a)$ denote the solution of (3) at time $t \geq 0$ for the initial condition $x(0) = a$. Since the state-variables correspond to normalized occupancy levels, we always assume that a belongs to the closed n-dimensional unit cube

$$C^n := \{x \in \mathbb{R}^n : x_i \in [0,1], i = 1, \ldots, n\}.$$

It has been shown in [34] that if $a \in C^n$ then $x(t,a) \in C^n$ for all $t \geq 0$, that is, C^n is an invariant set of the dynamics. Let $\text{int}(C^n)$ denote the interior of C^n, and let ∂C^n denote the boundary of C^n. Ref. [34] has also shown that the RFM is a *tridiagonal cooperative dynamical system* [57], and that (3) admits a *unique* steady-state point $e(\lambda_0, \ldots, \lambda_n) \in \text{int}(C^n)$ that is globally asymptotically stable, that is, $\lim_{t\to\infty} x(t,a) = e$ for all $a \in C^n$ (see also [33]). This means that the ribosome

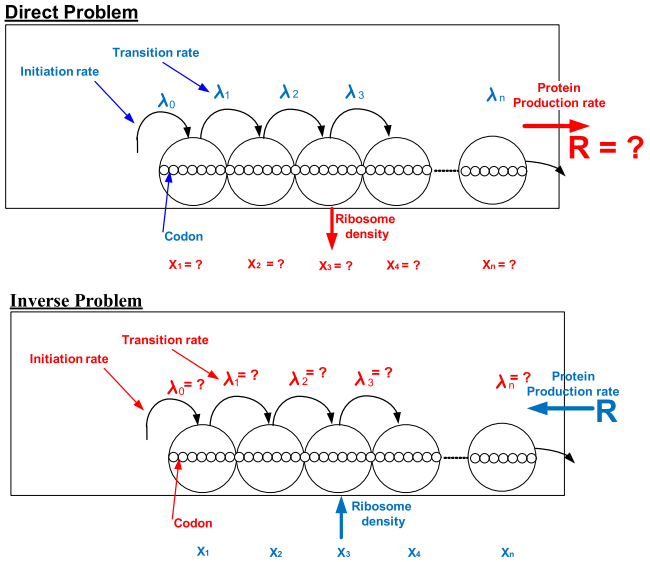


Fig. 2. Upper part: Previous studies considered the direct problem: given the RFM parameters, i.e., the set of transition rates λ_is, analyze the dynamics of the RFM ribosome densities x_is, and the production rate R. Lower part: Here we consider the inverse problem: Given a desired profile of ribosomal densities x_i, $i = 1, \ldots, n$, and a desired production rate R, find the rates that steer the dynamics to this profile.

density profile always converges to a steady-state profile that depends on the rates, but not on the initial condition. In particular, the production rate $R(t) = \lambda_n x_n(t)$ converges to a steady-state value

$$R := \lambda_n e_n. \tag{5}$$

At steady-state (i.e, for $x = e$), the left-hand side of all the equations in (3) is zero, so

$$\begin{aligned} \lambda_0(1 - e_1) &= \lambda_1 e_1(1 - e_2) \\ &= \lambda_2 e_2(1 - e_3) \\ &\vdots \\ &= \lambda_{n-1} e_{n-1}(1 - e_n) \\ &= \lambda_n e_n \\ &= R. \end{aligned} \tag{6}$$

This yields

$$R = \lambda_i e_i(1 - e_{i+1}), \quad i = 0, \ldots, n, \tag{7}$$

where $e_0 := 1$ and $e_{n+1} := 0$.

Remark 1. One may view (6) as a mapping from the rates $[\lambda_0, \ldots, \lambda_n]'$ to the steady-state density profile and production rate $[e_1 \ldots e_n \quad R]'$. For the purposes of this paper, it is important to note that this mapping is *invertible*. Indeed, Eq. (7) implies that given a desired density profile and production rate $[e_1 \ldots e_n \quad R]' \in (0, 1)^n \times \mathbb{R}_{++}$ one can immediately determine the transition rates that yield this profile at steady-state, namely,

$$\lambda_i = \frac{R}{e_i(1 - e_{i+1})}, \quad i = 0, \ldots, n. \tag{8}$$

Note that (8) implies that λ_i increases with R and e_{i+1}, and decreases with e_i. This is intuitive, as a larger λ_i implies a larger rate of ribosome flow from site i to site $i + 1$, as well as an increase in the steady-state production rate [43]. Thus, given a desired profile with larger R and e_{i+1}, and a smaller e_i, the required transition rates include a larger value for λ_i.

From a biophysical point of view, this means that if there are no constraints on the transition rates then we can engineer any desired density profile together with a desired production rate. More importantly, this provides an explicit expression for the needed rates. In addition to applications in functional genomics and molecular evolution, the observation in Remark 1 is also related to problems in synthetic biology where the goal is to re-engineer the mRNA molecule so as to obtain a desired density profile and production rate (see Fig. 2).

For more on the analysis of the RFM using tools from systems and control theory, see [35], [43], [44], [47], [68], [69]. The RFM models translation on a single isolated mRNA molecule. A network of RFMs, interconnected through a common pool of "free" ribosomes has been used to model simultaneous translation of several mRNA molecules while competing for the available ribosomes [48] (see also [1] for some related ideas).

It is important to mention that it has been shown in [49] that the correlation between the production rates based on modeling using RFM and using TASEP over all *S. cerevisiae* endogenous genes is 0.96, that the RFM agrees well with biological measurements of ribosome densities, and that the RFM predictions correlate well (correlations up to 0.6) with protein levels in various organisms (e.g., *E. coli*, *S. pombe*, *S. cerevisiae*). More recent results [16] show that a certain version of the RFM predicts well the density of RNA polymerases (RNAPs) during transcription. Given the high levels of bias related to the state of the art measurements of gene expression and the inherent noise in intracellular biological processes (see e.g., [13], [26]), these are very high correlations that demonstrate the relevance of the RFM in this context.

In this paper, we analyze the regulation of translation using the RFM. To do this, we first introduce two generalizations of the RFM into a control system.

2.1 The Controlled RFM

2.1.1 State- and Output-Controllability

Assume that every λ_i can be controlled independently. Thus, we replace every λ_i in the RFM by a function $u_i(t) : \mathbb{R}_+ \to \mathbb{R}_+$. The set of admissible controls $\mathbb{U}$ includes all the functions that are measurable, bounded, and take non-negative values for all $t \geq 0$. In the context of translation, manipulating the $u_i(t)$s corresponds to dynamically varying translation factors that regulate the initiation, elongation, and exit rates along the mRNA molecule. Note that we may view this as a networked control system: each state-variable represents an agent, the graph describing the agents interaction is a simple directed path, and the u_is control the strength of the graph edges. However, the dynamics of each agent is nonlinear.

The problem we consider is whether it is possible, using the $n + 1$ control functions, to steer x and R from any initial

condition to any desired conditions $x^f \in \text{int}(C^n)$ and $R^f \in \mathbb{R}_{++}$ in finite time, and if so, to determine appropriate controls.

Of course, independently controlling all the transition rates may be difficult to do in practice, so we also consider another controlled version of the RFM.

2.1.2 Output-Controllability

Assume that a subset of m rates $\lambda_{j_1}, \ldots, \lambda_{j_m}$, with $1 \leq m \leq n+1$, can be *jointly* controlled, i.e., all these rates can be replaced by a common, scalar, non-negative control function $u(t)$. This models the case where a single factor jointly controls one or more transition rates.

For example in an RFM with length $n = 3$, assume that the rates λ_1 and λ_2 can be replaced by a common, scalar, non-negative control function $u(t)$. The resulting model is

$$\begin{aligned}\dot{x}_1 &= \lambda_0(1 - x_1) - ux_1(1 - x_2),\\ \dot{x}_2 &= ux_1(1 - x_2) - ux_2(1 - x_3),\\ \dot{x}_3 &= ux_2(1 - x_3) - \lambda_3 x_3.\end{aligned}$$

This scenario is biologically relevant since the exact same codon may appear in multiple places along the transcript, and since the same tRNA species may moreover be involved in the decoding of more than a single codon through wobble pairing. Thus, regulating the abundance of a single tRNA molecule would typically have a simultaneous effect on transition rates at multiple positions along the mRNA transcript. In the context of this problem, we are interested in using $u(t)$ to steer only the production rate R to a desired value R^f in finite time. Specifically, the problem that we consider is whether it is possible to use u to steer R from any initial condition to any feasible value and, if so, to determine a suitable control u. Of course, the set of feasible values is determined by the other, $n + 1 - m$ fixed transition rates.

We show that both control problems described above are *controllable*. In other words, the control authority is always powerful enough to obtain any feasible desired density profile and/or production rate. This is a primarily theoretical result. However, we also show that there exist positive and constant controls that asymptotically steer the controlled RFM to the desired densities/production rate. In the problem of controlling all the rates, these constant values are given in a simple and closed-form expression. In the second control problem, this constant value can be easily found numerically using a simple line search algorithm.

We now discuss the biological relevance of these control problems. Understanding and manipulating the mRNA translation rate is related to numerous biomedical disciplines including human health, evolution, genetics, biotechnology, and more [2], [3], [4], [27], [29], [37], [50], [63], [66]. Controlling the entire ribosomal density profile, and not only the translation rate, by manipulating the transition rates is also a fundamental problem as it is known that the density profile along the mRNA molecule is important for various intracellular phenomena. For example, it was shown that the density and induced speed of ribosome flow along the mRNA affect co-translational folding of the protein. If the density and the induced flow speed of the ribosomes is inappropriate then the protein may misfold leading to a nonfunctional protein (see, for example, [27], [29], [40], [70]). In addition, it was suggested that the density of ribosomes affects mRNA degradation: a higher ribosome density is related to lower efficiency of mRNA degradation and longer half life [11], [14], [17], [41]. Furthermore, ribosome density is directly related to ribosomal collisions and translation abortion [2], [18], [61], [63], [73]: a higher density increases the probability of collisions and may lead to abortions and thus the production of truncated and potentially deleterious proteins. Finally, ribosome density is strongly correlated with ribosome allocation: a higher density of ribosomes on the mRNA decreases the pool of free ribosomes, the initiation rate in other mRNA molecules, and thus the organism growth rate and fitness [2], [18], [61], [63], [73].

Our results suggest that these important issues can be addressed using a combination of mathematical, computational, and experimental approaches. Our results also provide an initial but explicit solution to the problem of controlling the augmented profile. While the model and problems are relatively simple, they may still provide a reasonable approximation to the biological solution in some cases. They may also be used as a starting point for addressing and solving similar problems in more comprehensive models of translation.

The next section describes our main results. Readers who are not familiar with controllability analysis may consult Appendix A for a quick review of this topic.

3 Main Results

As noted above, we consider two control problems for the RFM. We now detail their exact mathematical formulation, and then present our main results.

3.1 Controlling the State and the Output

Let $\Omega := C^n \times \mathbb{R}_+$. Assume first that all the $n+1$ transition rates can be controlled. The control is then $u(t) = [\,u_0(t), \ldots, u_n(t)\,]'$ and the dynamics of the controlled RFM with output $R(t)$ is described by

$$\begin{aligned}\dot{x}_i(t) &= u_{i-1}(t)x_{i-1}(t)(1 - x_i(t)) - u_i(t)x_i(t)(1 - x_{i+1}(t)),\\ &\qquad i = 1, \ldots, n,\\ R(t) &= u_n(t)x_n(t).\end{aligned} \tag{9}$$

We define the admissible set $\mathbb{U}$ as the set of measurable and bounded controls taking values in $\mathbb{R}_+^{n+1}$ for all time t.

Problem 1. *Given arbitrary $x^s, x^f \in \text{int}(C^n)$ and $R^s, R^f \in \mathbb{R}_{++}$, does there always exist a time $T \geq 0$ and a control $u \in \mathbb{U}$ such that $x(T, u, x^s) = x^f$ and $R(T, u, R^s) = R^f$? If so, determine such a control.*

We can now state our first main result. Recall that all the proofs are placed in Appendix B.

Theorem 1. *The controlled RFM (9) is state- and output-controllable on $\text{int}(\Omega)$. Furthermore, for any $x^f = \begin{bmatrix} x_1^f \ldots x_n^f \end{bmatrix}' \in \text{int}(C^n)$ and $R^f \in \mathbb{R}_{++}$, define $v \in \mathbb{R}_{++}^{n+1}$ by*

$$v_i := \frac{R^f}{x_i^f(1 - x_{i+1}^f)}, \quad i = 0, \ldots, n, \tag{10}$$

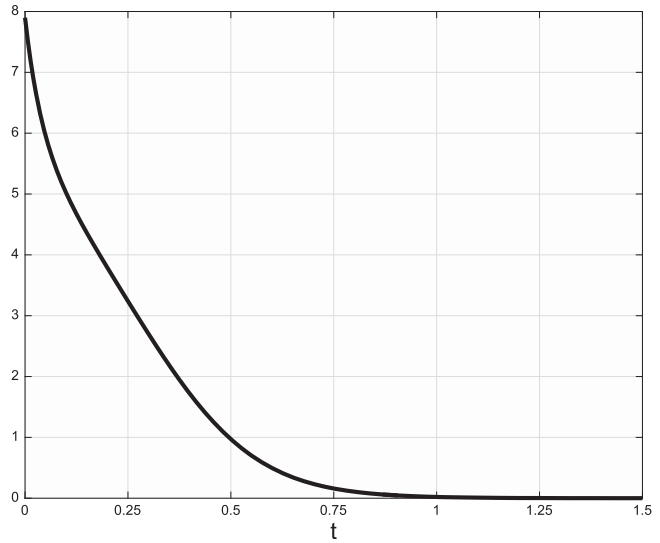


Fig. 3. The error $|x(t)-x^f|_1+|R(t)-R^f|_1$ as a function of t in Example 1.

where $x_0^f := 1$ *and* $x_{n+1}^f := 0$. *Then for any* $x^s \in C^n$ *and any* $R^s \in \mathbb{R}_+$ *applying the constant control* $u(t) \equiv v$ *in (10) yields*

$$\lim_{t\to\infty} x(t,u,x^s) = x^f, \quad \lim_{t\to\infty} R(t,u,R^s) = R^f. \tag{11}$$

This means that the control is "powerful" enough to steer the system, in *finite time*, from any initial augmented profile to any desired final augmented profile. It also provides a simple closed-form solution for a control that *asymptotically* steers the system to x^f and R^f from any initial condition. In other words, it practically solves the control synthesis problem.

An important property of v is that it does not depend on the initial values x^s and R^s, but only on the desired augmented profile (x^f, R^f). This is important as measuring x^s, that is, the initial ribosomal profile along the mRNA, may be difficult due to the current limitations in measuring ribosome densities (see, for example, [8], [10], [13]).

Example 1. Consider the controlled RFM with dimension $n = 5$. Suppose that we would like to steer the ribosomal density profile along the mRNA molecule to $[0.8 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1]'$, and the production rate to 1.5. The profile here is motivated by the fact that low ribosome abundance at the beginning of the ORF reduces ribosome "traffic jams" that may lead to ribosome drop off. Setting $x^f = [0.8 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1]'$, $R^f = 1.5$, and applying (10) yields

$$v = [15/2 \quad 25/12 \quad 50/3 \quad 50/3 \quad 50/3 \quad 15]'.$$

Fig. 3 depicts the error $|x(t,u,x^s)-x^f|_1+|R(t,u,R^s)-R^f|_1$ (where $|z|_1$ denotes the L_1 norm of the vector z) for the initial conditions $x^s = [0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5]'$, $R^s = 0.5$, and the control $u(t) \equiv v$. It may be observed that the error decays at an exponential rate to zero. Thus, this control steers the system arbitrarily close to the desired final density profile x^f and production rate R^f.

Example 1 suggests that the explicit constant control in Theorem 1 provides a good practical solution to Problem 1.

3.2 Controlling the Output

Pick an arbitrary set of indexes $\Theta \subseteq \{0,\dots,n\}$, and let $m := |\Theta|$. Replace every λ_i, $i \in \Theta$, in the RFM by a common, scalar control $u(t)$. Pick $c > 0$, and assume that $u(t) \in [0,c]$, for all $t \geq 0$, i.e., the set of admissible controls $\mathbb{U}$ is the set of measurable scalar functions taking values in $[0,c]$ for all $t \geq 0$. As noted above, this formulation represents a biologically relevant scenario, as we assume that several translation rates are controlled by the same control, and also that the allowed control action is bounded by the value c.

Our goal is to use the scalar control to regulate the production rate $R(t)$, i.e., the output. Of course, not every value of $R(t)$ is possible, because of the non-regulated, fixed transition rates. One can in principle define the reachable set of $R(t)$ based on the fact that the state trajectories evolve on C^n. For example, if $n \notin \Theta$ then $R(t) = \lambda_n x_n(t)$ implies that one can define the reachable set as $[0,\lambda_n]$. However, this definition is not really relevant. Indeed, assume that some rate λ_k, with $k \notin \Theta$, is much smaller than all the other rates and also much smaller than c. Then regardless of the specific control used it is clear that after some time $R(t)$ will also be small, as λ_k will be the limiting factor, and so after some time it will become impossible to steer the production rate to every desired value in the set $[0,\lambda_n]$.

We define a more meaningful reachable set for the production rate as follows. Let $\bar{\lambda} \in \mathbb{R}_{++}^{n+1-m}$ denote the set of fixed transition rates. For every time $T \geq 0$ and every initial condition $x_0 \in C^n$, let $\Omega(\bar{\lambda},\Theta,c,T,x_0) \subset \mathbb{R}_+$ denote the set of production rates that can be attained at some time $t \geq T$ with $x(0) = x_0$. Define the *large-time reachable set* of R as

$$\Omega(\bar{\lambda},\Theta,c,x_0) := \cap_{T\geq 0}\Omega(\bar{\lambda},\Theta,c,T,x_0).$$

Although the RFM is a nonlinear model, this set can be characterized explicitly. To derive this characterization, we introduce more notation. First, define a vector $q \in \mathbb{R}^{n+1}$ by

$$q_i := \begin{cases} c, & i \in \Theta, \\ \lambda_i, & \text{otherwise.} \end{cases}$$

For example, for $\Theta = \{1,2,n\}$, $q = [\lambda_0, c, c, \lambda_3, \dots, \lambda_{n-1}, c]'$.

Also, for $\ell_0,\dots,\ell_n > 0$ define a $(n+2)\times(n+2)$ symmetric, tridiagonal, and componentwise nonnegative matrix $A = A(\ell_0,\dots,\ell_n)$ by

$$A := \begin{bmatrix} 0 & \ell_0^{-1/2} & 0 & 0 & \dots & 0 & 0 \\ \ell_0^{-1/2} & 0 & \ell_1^{-1/2} & 0 & \dots & 0 & 0 \\ 0 & \ell_1^{-1/2} & 0 & \ell_2^{-1/2} & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & \dots & \ell_{n-1}^{-1/2} & 0 & \ell_n^{-1/2} \\ 0 & 0 & 0 & \dots & 0 & \ell_n^{-1/2} & 0 \end{bmatrix}, \tag{12}$$

and let $\zeta_{MAX}(A)$ denote the maximal eigenvalue of A.[1] The next result uses the linear-algebraic representation of the steady-state production rate in the RFM derived in [43].

1. It is clear that the eigenvalues are real as A is symmetric. Since A is also nonnegative and irreducible the eigenvalues are distinct.

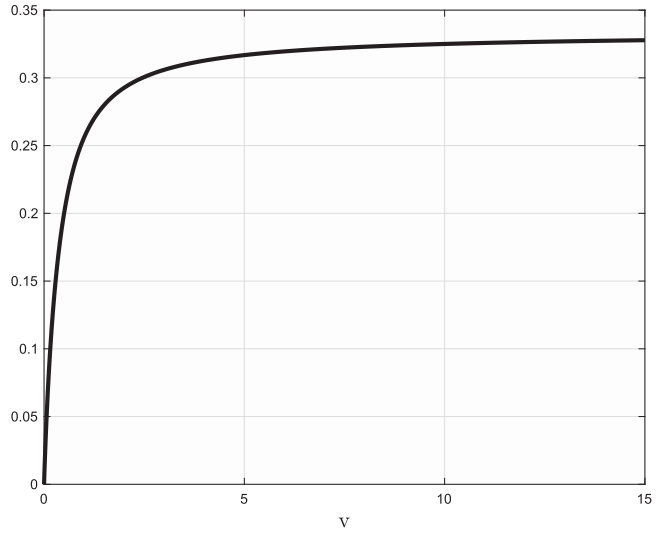


Fig. 4. Maximal steady-state production rate $(\zeta_{MAX}(A(1,\ 1/2, v, 3, v,\ 1/2)))^{-2}$ for $v \in [0, 15]$.

Proposition 1. *For any* $x_0 \in C^n$,

$$\Omega(\bar{\lambda}, \Theta, c) = [0, M], \tag{13}$$

where $M := (\zeta_{MAX}(A(q_0, \ldots, q_n)))^{-2}$.

Note that (13) implies that $\Omega(\bar{\lambda}, \Theta, c)$ does not depend on x_0, but only on the vector q.

Remark 2. Denote the indexes in Θ by $j_1, \ldots, j_m$. Consider the case $c \to \infty$. Then $c^{-1/2} \to 0$, so the largest eigenvalue of the matrix $A(q_0, \ldots, q_n)$ tends to

$$\max\{\zeta_{MAX}(Q_0), \ldots, \zeta_{MAX}(Q_m)\},$$

where

$$\begin{aligned} Q_0 &:= A(\lambda_0, \ldots, \lambda_{j_1-1}), \\ Q_k &:= A(\lambda_{j_k+1}, \ldots, \lambda_{j_{k+1}-1}), \quad k = 1, \ldots, m-1, \\ Q_m &:= A(\lambda_{j_m+1}, \ldots, \lambda_n), \end{aligned} \tag{14}$$

with $\zeta_{MAX}(B) := 0$ if B is an empty matrix. Thus, in this case

$$M = \min\{(\zeta_{MAX}(Q_0))^{-2}, \ldots, (\zeta_{MAX}(Q_m))^{-2}\}, \tag{15}$$

where 0^{-2} is defined as ∞. In other words, when the maximal control value of the controlled transition rates goes to infinity, the maximal possible steady-state production rate will be the minimum of the steady-state production rates of several RFMs: the first with rates $\lambda_0, \ldots, \lambda_{j_1-1}$, the second with rates $\lambda_{j_1+1}, \ldots, \lambda_{j_2-1}$, and so on, with the last RFM with rates $\lambda_{j_m+1}, \ldots, \lambda_n$. This demonstrates how in this case the other, fixed rates, being the limiting factors, determine the feasible set for the production rate.

From the biological point of view this means that if the transition rates along some regions of the mRNA are very high (and thus not rate limiting) the production rate will depend only on the transition rates before and after this region, as these include the rate limiting factor. Also, the large-time reachable set for the production rate will be constrained by the rate limiting transition rates.

Example 2. Consider a controlled RFM with length $n = 5$, $\Theta = \{2, 4\}$, and fixed rates

$$\lambda_0 = 1,\ \lambda_1 = 1/2,\ \lambda_3 = 3,\ \lambda_5 = 1/2. \tag{16}$$

In other words, λ_2 and λ_4 are both replaced by the scalar control $u(t)$. Suppose that the admissible set $\mathbb{U}$ is the set of functions taking values in $[0, c]$, with $c = 15$. Fig. 4 depicts $(\zeta_{MAX}(A(1, 1/2, v, 3, v, 1/2)))^{-2}$ for $v \in [0, 15]$. It may be seen that this is a strictly increasing function of v. A calculation yields (all numbers are to four digit accuracy)

$$(\zeta_{MAX}(A(1, 1/2, 15, 3, 15, 1/2)))^{-2} = 0.3278,$$

so $\Omega = [0, 0.3278]$.

Note that if we take $c \to \infty$ then (14) yields

$$\begin{aligned} Q_0 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & (1/2)^{-1/2} \\ 0 & (1/2)^{-1/2} & 0 \end{bmatrix}, \\ Q_1 &= \begin{bmatrix} 0 & 3^{-1/2} \\ 3^{-1/2} & 0 \end{bmatrix}, \\ Q_2 &= \begin{bmatrix} 0 & (1/2)^{-1/2} \\ (1/2)^{-1/2} & 0 \end{bmatrix}, \end{aligned}$$

and so (15) yields

$$\begin{aligned} &\min\{(\zeta_{MAX}(Q_0))^{-2}, (\zeta_{MAX}(Q_1))^{-2}, (\zeta_{MAX}(Q_2))^{-2}\} \\ &\quad = \min\{1/3, 3, 1/2\} \\ &\quad = 1/3. \end{aligned}$$

The next result considers controlling the output to a desired value in $\Omega(\bar{\lambda}, \Theta, c)$.

Proposition 2. *The controlled RFM with one or more rates replaced by a common scalar control function* $u(t)$ *is output-controllable in* $\mathrm{int}(\Omega(\bar{\lambda}, \Theta, c))$. *Furthermore, for any* $R^f \in \mathrm{int}(\Omega(\bar{\lambda}, \Theta, c))$ *there exists a value* $v \in [0, c]$ *such that the constant control* $u(t) \equiv v$ *yields* $\lim_{t\to\infty} R(t) = R^f$.

This means that jointly regulating one or more transition rates with a common scalar control function $u(t)$ is still "powerful" enough to steer the production rate from any initial value to any desired final value $R^f \in \mathrm{int}(\Omega)$ in finite time. Furthermore, the controlled RFM is asymptotically controllable in Ω, even when $\mathbb{U}$ is restricted to constant controls only. Since $\zeta_{MAX}(A(\ell_0, \ldots, \ell_n))$ is a strictly decreasing function of every ℓ_i, finding the constant value v that asymptotically steers the system to a desired value $R^f \in \mathrm{int}(\Omega)$ can be easily solved numerically using a simple line search. The next example demonstrates this.

Example 3. Consider again the controlled RFM in Example 2. Recall that the admissible set $\mathbb{U}$ is the set of functions taking values in $[0, c]$, with $c = 15$. We already know that in this case $\Omega = [0, 0.3278]$. Assume that our goal is to asymptotically steer the production rate to, say, $R^f = 0.3$. A simple line search shows that the corresponding constant control value is $v = 2.4534$ (see also Fig. 4).

3.3 Sensitivity Analysis

In practice, the applied controls are never exactly equal to the desired values and therefore it is important to understand the effect of small perturbations in the control values on the desired augmented profile. Since we are basically considering constant controls, it is enough to study the sensitivity of the steady-state density profile of the RFM to small changes in the λ_is. (The sensitivity of the steady-state production rate R with respect to the λ_is has been studied in [44].)

Proposition 3. *Consider the RFM with dimension n, and let $e := [\, e_1 \dots e_n \,]'$ denote the corresponding steady-state point in $\mathrm{int}(C^n)$. Pick an index $i \in \{0, \dots, n\}$. Then $\frac{\partial}{\partial \lambda_i} e_k$ exists for all k, and*

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} e_k &< 0, \quad \textit{for all } k \leq i, \\ \frac{\partial}{\partial \lambda_i} e_k &> 0, \quad \textit{for all } k > i. \end{aligned} \tag{17}$$

Thus, increasing λ_i decreases [increases] the steady-state densities in sites $1, \dots, i$ [sites $i+1, \dots, n$]. This is reasonable, as increasing λ_i increases the transition rate from site i to site $i+1$ (see also [48] for some related considerations).

Example 4. Recall from Example 1 that for the RFM with $n = 5$ the control

$$u(t) \equiv [\, 15/2 \quad 25/12 \quad 50/3 \quad 50/3 \quad 50/3 \quad 15 \,]',$$

yields the steady-state augmented profile

$$[\, e \quad R \,]' = [\, 0.8 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 1.5 \,]'. \tag{18}$$

Let $\tilde{u}(t) \equiv [\, 15/2 \quad 25/12 \quad (50/3) + \varepsilon \quad 50/3 \quad 50/3 \quad 15 \,]'$, with $\varepsilon := 0.2$ i.e., the same transition rates as before, but with ε added to λ_2. Using (6) shows that $\tilde{u}$ yields the steady-state augmented profile

$$[\, \tilde{e} \quad \tilde{R} \,]' = [\, 0.7998 \quad 0.0989 \quad 0.1001 \quad 0.1001 \quad 0.1001 \quad 1.5013 \,]',$$

(all numbers are to four digit accuracy). Comparing this to (18) shows that the steady-state values at sites 1,2 decreased, and those at sites 3,4,5 increased.

4 Discussion

Regulating the ribosomal density profile along the mRNA molecule, and not only the protein production rate, is an important problem in evolutionary biology, biotechnology, and synthetic biology because this density profile affects various fundamental intracellular processes including mRNA degradation, protein folding, ribosomal allocation and abortion, and more (see, for example, [17], [27], [29], [40], [63], [70]). It seems that there are still considerable gaps in our understanding of how the density profile is regulated, and how it can be re-engineered. In this paper, we addressed this issue by analyzing a mathematical model for ribosome flow, the RFM, using tools from nonlinear control theory.

Our results indicate that if we are able to control all the transition rates along the different parts of the mRNA then we can steer the system to any desired ribosomal density profile, and we provide a closed-form expression for a constant control vector that achieves this asymptotically.

Also, jointly controlling one or more transition rates using a common scalar control allows to steer the protein production rate to any desired value within a feasible range that is determined by the other, fixed transition rates. A simple line search algorithm can be used to derive a constant control value that achieves this asymptotically. This case models scenarios where for example the abundance of a specific loaded tRNA molecule is regulated. Indeed, regulating the abundance of a certain tRNA molecule should simultaneously affect the translation rate at all the positions along the mRNA with corresponding codons. Typically, a certain codon may repeat at dozens, or even hundreds of locations along one mRNA molecule.

Our results are based on the RFM that, as any mathematical model, is a simplification of (the biological) reality. For example, the RFM does not encapsulate some of the complex interactions between the transcript features and translation (see, e.g., [51], [62], [63]). Nevertheless, using the RFM allows one to pose the controllability and control synthesis problems in a well-structured way, and study them rigorously using tools from systems and control theory.

We believe that our analytical results may lead to new biological insights and suggest novel and interesting biological experiments. For example, it has been suggested that a higher ribosome density contributes to a higher mRNA half life in *S. cerevisiae* [17]. However, it is difficult to determine if the correlation is due to a larger abundance of ribosomes along the entire coding region or maybe only the ribosome density at the 5′end of the coding region is relevant. It is also possible that this relation is due to a higher number of pre-initiation complexes at the 5′UTR (that contribute to a higher initiation rate). Specifically, it is possible that only higher pre-initiation density or ribosome density at the 5′end is important since in some cases the degradation starts from this region. Both factors are expected to correlate with higher ribosome density along the entire coding region, and a natural question is how can we design an experiment that can separate between the two possible explanations?

The results reported here suggest that we can design a synthetic library (that can be studied in-vitro and/or in-vivo) with different strains that have different initiation rates, but identical ribosome densities along the coding regions, or strains with different levels of ribosome densities at the first codons (or any other segment) of the coding regions, but similar ribosome densities in the rest of the coding region. Using such libraries may help in understanding exactly which factor contributes to the higher mRNA half life.

Regulating transition rates can also affect the folding of the protein. Indeed, it was suggest in [38] that synonymous codons substitutions, that change the corresponding transition rates, may switch some protein domains between post-translationally and co-translationally folding.

We believe that the results reported in this study may also contribute towards a better understanding of the molecular evolution of translation. Since usually a change in a transition rate is related to a mutation/change in the mRNA codons composition, obtaining a desired ribosomal density profile and production rate involves introducing changes in the nucleotide composition of the transcript.

Thus, an important future study should combine controllability analysis with models of molecular evolution.

Other topics for further research include the following. First, from the biological point of view a relevant scenario is when some of the transition rates can be controlled, but each rate can take values in a discrete set of possible values only. Indeed, the admissible rates are limited by factors such as the concentrations of initiation and elongation factors, and the biophysical properties of the ribosome, mRNA, and translation factors. In this case, it is clear that we cannot obtain any desired density profile, and an interesting problem may be to determine the rate values that yield the "best" approximation for a given desired profile. This requires a biologically relevant definition of this best approximation, i.e., a measure of distance between two density profiles that is biologically relevant.

Second, as noted above, the RFM is a mean-field approximation of TASEP. Our results naturally raise the question of whether TASEP is controllable (in some stochastic sense). It is also interesting to examine if the analytical results obtained for the RFM can be used to synthesize suitable hopping rates for the stochastic TASEP model. In other words, suppose that we are given a desired profile P for the RFM, and determine the corresponding constant rates v_is using (10). Does using these rates (perhaps after some normalization) as the TASEP hopping rates yield the steady-state profile P in TASEP as well?

Finally, TASEP has been used to model and analyze many other applications, for example, traffic flow. The RFM can also be used to study these applications, and controllability and control synthesis may be important here as well. For example, a natural question is can the density along a traffic lane be steered to any arbitrary profile by regulating speed signs along different sections of the lane?

APPENDIX
APPENDIX A: REVIEW OF CONTROLLABILITY

Controllability is a fundamental property of control systems, but it is not necessarily well-known outside of the systems and control community. For the sake of completeness, we briefly review this topic here. For more details, see e.g., [58].

Consider the control system

$$\begin{aligned} \dot{x} &= f(x,u), \\ y &= h(x,u), \end{aligned} \tag{19}$$

where $x : \mathbb{R}_+ \to \mathbb{R}^n$ is the state vector, $u : \mathbb{R}_+ \to \mathbb{R}^m$ is the control, and $y : \mathbb{R}_+ \to \mathbb{R}^k$ is the output. Let $\mathbb{U}$ denote the set of admissible controls. Assume that the trajectories of this system evolve on a state space $\Omega \subseteq \mathbb{R}^n$. Given an initial condition $a \in \Omega$ and a desired final condition $b \in \Omega$, a natural control problem is: find a time $T \geq 0$, and an admissible control $u : [0, T] \to \mathbb{R}^m$ such that

$$x(T, u, a) = b.$$

In other words, u steers the system from a to b in time T. Of course, such a control may not always exist. This leads to the following definition.

Definition A.1. *The system (19) is said to be* state-controllable *on Ω if for any $a, b \in \Omega$ there exist a time $T \geq 0$, and a control $u \in \mathbb{U}$ such that $x(T, u, a) = b$.*

Sometimes it is enough to steer only the output to a desired condition. This leads to the following definition.

Definition A.2. *The system (21) is said to be* output-controllable *on some set $\Psi \subseteq \mathbb{R}^k$ if for any $p, q \in \Psi$ there exist a time $T \geq 0$, and a control $u \in \mathbb{U}$ that steers the output from $y(0) = p$ to $y(T) = q$.*

Controllability is thus a theoretical property, but it is important in many applications, as it implies that the problem of determining a suitable control, i.e., the *control synthesis problem*, always admits a solution. From here on we focus on state-controllability. The notions for output-controlability are analogous.

Another useful notion, that is weaker than controllability, is called asymptotic controllability.

Definition A.3. *System (19) is said to be* asymptotically state-controllable *on Ω if for any $a, b \in \Omega$ there exists a control $u \in \mathbb{U}$ such that*

$$\lim_{t \to \infty} x(t, u, a) = b.$$

Note that this implies that for any neighborhood V of b, there exists a time $T_s \geq 0$, and a control $u_s \in \mathbb{U}$ such that $x(T_s, u_s, a) \in V$.

For nonlinear control systems, analyzing controllability or asymptotic controllability is not trivial. There exists a weaker theoretical notion that can be analyzed effectively using Lie-algebraic techniques. For $a \in \Omega$, define the *reachable set from a* by

$$RS(a) := \{x(t, u, a) : t \geq 0,\ u \in \mathbb{U}\}.$$

In other words, $RS(a)$ is the set of all states that can be reached at some time $t \geq 0$ starting from $x(0) = a$. The system (19) is said to be *accessible* from a if the set $RS(a)$ has a non empty interior. In other words, the control is powerful enough to allow steering the trajectories emanating from a to a "full set" of directions.

Example A.1. Consider the scalar system $\dot{x} = u$, with $\Omega = \mathbb{R}$. Let $\mathbb{U}$ be the set of measurable functions taking non-negative values for all time t. Pick $a \in \Omega$. Then $RS(a) = [a, \infty)$, so the systen is accessible from a. However, the system is not controllable on Ω, as there does not exist any control $u \in \mathbb{U}$ that steers a to a point b with $b < a$.

Our results for the controlled RFM are based on proving that it is asymptotically state-controllable, using *constant* controls, and combining this with a Lie-algebraic sufficient condition for accessibility to deduce state-controllability.

To describe a sufficient condition for accessibility, consider the *control affine system*

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x) u_i, \tag{20}$$

and assume that $0 \in \mathbb{U}$. For two vector fields $f, g : \mathbb{R}^n \to \mathbb{R}^n$, let $[f,g] := \frac{\partial g}{\partial x} f - \frac{\partial f}{\partial x} g$. This is another vector field called the *Lie-bracket of f and g*. For example, if $f(x) = Ax$ and $g(x) = Bx$ then $[f,g](x) = (BA - AB)x$. It is useful to introduce a notation for iterated Lie brackets. These can be defined inductively by letting $\mathrm{ad}_f^0 g := g$, $\mathrm{ad}_f^1 g := [f,g]$, and $\mathrm{ad}_f^k g := [f, \mathrm{ad}_f^{k-1} g]$ for any integer $k \geq 1$.

The *Lie algebra* A_{LA} associated with (20) is the linear subspace that is generated by $\{f, g_1, \ldots, g_m\}$ and is closed under the Lie bracket operation. Let

$$A_{LA}(x_0) := \{p(x_0) : p \in A_{LA}\}.$$

Roughly speaking, it can be shown that if small-time solutions of (22) emanating from a point x_0 and corresponding to piecewise constant controls "cover" a k-dimensional set, with $k \leq n$, then $A_{LA}(x_0) = \mathbb{R}^k$. This yields the following sufficient condition for accessibility.

Theorem A.1. *If $A_{LA}(x_0) = \mathbb{R}^n$ at some point x_0 then (20) is accessible from x_0.*

The next result applies Theorem A.1 to analyze accessibility in the RFM when either the entry rate or exit rate is replaced by a control.

Fact A.1. *Consider the n-dimensional RFM with a single rate λ_i replaced by a scalar control $u(t)$. If $i = 0$ or $i = n$ then the control system is accessible from any point $x \in \mathrm{int}(C^n)$.*

Proof of Fact A.1. Consider the controlled RFM obtained by replacing λ_0 by $u(t)$, leaving the other rates as strictly positive constants. Let $z_0(x) := \lambda_0(1 - x_1)$, $z_j(x) := \lambda_j x_j (1 - x_{j+1})$, for $j = 1, \ldots, n-1$, and $z_n(x) := \lambda_n x_n$. The controlled RFM satisfies

$$\dot{x} = f(x) + g(x)u, \tag{21}$$

where $f := [\, -z_1 \quad z_1 - z_2 \quad z_2 - z_3 \ldots z_{n-1} - z_n \,]'$, and $g := [\, 1 - x_1 \quad 0 \ldots 0 \,]'$. Let $p^k(x) := (\mathrm{ad}_f^k g)(x)$. A calculation shows that for all $k \in \{0, \ldots, n-1\}$,

$$p^k = \left[\, p_1^k \ldots p_k^k \quad p_{k+1}^k \quad 0 \ldots 0 \,\right]',$$

with $p_{k+1}^k = (-1)^k \prod_{j=1}^{k+1}(1 - x_j) \prod_{\ell=1}^{k} \lambda_\ell$. Note that $p_{k+1}^k \neq 0$ for all $x \in \mathrm{int}(C^n)$, so the n vector fields $p^0, \ldots, p^{n-1}$ are linearly independent, and thus span $\mathbb{R}^n$. Thus, the controlled RFM is accessible from any $x \in \mathrm{int}(C^n)$.

Now consider the case where λ_n is replaced by a control $u(t)$. For $j = 1, \ldots, n$, let $q_j(t) := 1 - x_{n+1-j}(t)$. Then

$$\begin{aligned}
\dot{q}_1 &= (1 - q_1)u - \lambda_{n-1} q_1 (1 - q_2), \\
\dot{q}_2 &= \lambda_{n-1} q_1 (1 - q_2) - \lambda_{n-2} q_2 (1 - q_3), \\
&\vdots \\
\dot{q}_n &= \lambda_1 q_{n-1} (1 - q_n) - \lambda_0 q_n.
\end{aligned}$$

This is a controlled RFM with the initiation rate replaced by a control $u(t)$. It follows from the analysis above that this control system is accessible in $\mathrm{int}(C^n)$, and this completes the proof. □

Another sufficient condition for accessibility is based on *linearizing* the control system around an equilibrium point. For our purposes, it is enough to state this condition for the control affine system (20) with $m = 1$, i.e., the system

$$\dot{x} = f(x) + g(x)u. \tag{22}$$

Theorem A.2 [58, Ch. 3]. *Suppose that $f(e) = 0$ and that $0 \in \mathrm{int}(\mathbb{U})$. Consider the linear control system*

$$\dot{z} = Az + ub,$$

where $A := \frac{\partial f}{\partial x}(e)$ and $b := g(e)$. If the $n \times n$ matrix $\left[\, b \quad Ab \quad \ldots A^{n-1} b \,\right]$ is invertible then (22) is accessible from some neighborhood of e.[2]

Example A.2. Consider the RFM with $n = 2$, i.e.,

$$\begin{aligned}
\dot{x}_1 &= \lambda_0(1 - x_1) - \lambda_1 x_1 (1 - x_2), \\
\dot{x}_2 &= \lambda_1 x_1 (1 - x_2) - \lambda_2 x_2,
\end{aligned} \tag{23}$$

with $\lambda_i > 0$. The steady-state point e of this system satisfies $\lambda_0(1 - e_1) = \lambda_1 e_1 (1 - e_2) = \lambda_2 e_2$. Suppose now that we can control the transition rate from site 1 to site 2. To study state-controllability in the neighborhood of e, consider the control system

$$\begin{aligned}
\dot{x}_1 &= \lambda_0(1 - x_1) - (\lambda_1 + u) x_1 (1 - x_2), \\
\dot{x}_2 &= (\lambda_1 + u) x_1 (1 - x_2) - \lambda_2 x_2,
\end{aligned} \tag{24}$$

where $\mathbb{U}$ is the set of measurable functions taking values in $[-\varepsilon, \varepsilon]$ for some sufficiently small $\varepsilon > 0$. This system is in the form (22) with $f(x) = [\, \lambda_0(1 - x_1) - \lambda_1 x_1 (1 - x_2) \; \lambda_1 x_1 (1 - x_2) - \lambda_2 x_2 \,]'$, and $g(x) = x_1(1 - x_2)[\, -1 \quad 1 \,]'$. Note that $f(e) = 0$. To apply Theorem A.2, calculate $A = \begin{bmatrix} -\lambda_0 - \lambda_1(1 - e_2) & \lambda_1 e_1 \\ \lambda_1(1 - e_2) & -\lambda_1 e_1 - \lambda_2 \end{bmatrix}$, $b = e_1(1 - e_2)[\, -1 \quad 1 \,]'$, and

$$[\, b \quad Ab \,] = e_1(1 - e_2) \begin{bmatrix} -1 & \lambda_0 + \lambda_1(1 - e_2) + \lambda_1 e_1 \\ 1 & -\lambda_1(1 - e_2) - \lambda_1 e_1 - \lambda_2 \end{bmatrix}.$$

Note that $\det([\, b \quad Ab \,]) = e_1^2 (1 - e_2)^2 (\lambda_2 - \lambda_0)$. Since $e \in \mathrm{int}(C^2)$, Theorem A.2 implies that if $\lambda_0 \neq \lambda_2$ then (24) is accessible in a neighborhood of e.

Now consider (24) with $\lambda_0 = \lambda_2$. Then $z := x_1 + x_2$ satisfies

$$\dot{z} = \lambda_0 (1 - z).$$

Thus, any trajectory with $x_1(0) + x_2(0) = 1$ satisfies $x_1(t) + x_2(t) \equiv 1$ for any control u, and this implies that in this case (24) is not accessible and not state-controllable on C^2.

Summarizing, in this case the condition in Theorem A.2 allows us to completely analyze the accessibility of (24).

This example may suggest that accessibility is lost when one of the *internal* (or elongation) rates λ_i, $i \in \{1, \ldots, n-1\}$,

2. In fact, the condition above guarantees a stronger property, called first-order local controllability, but for our purposes the more restricted statement in Theorem A.2 is enough.

is replaced by a control, at least for some values of the other rates. However, the next example shows that is not necessarily true.

Example A.3. Consider the RFM with $n = 3$, i.e.,

$$\begin{aligned}\dot{x}_1 &= \lambda_0(1-x_1) - \lambda_1 x_1(1-x_2),\\ \dot{x}_2 &= \lambda_1 x_1(1-x_2) - \lambda_2 x_2(1-x_3),\\ \dot{x}_3 &= \lambda_2 x_2(1-x_3) - \lambda_3 x_3,\end{aligned}$$

with $\lambda_i > 0$. Suppose that we can control the transition rate from sites 1 to 2, so we consider the control system

$$\begin{aligned}\dot{x}_1 &= \lambda_0(1-x_1) - x_1(1-x_2)u,\\ \dot{x}_2 &= x_1(1-x_2)u - \lambda_2 x_2(1-x_3),\\ \dot{x}_3 &= \lambda_2 x_2(1-x_3) - \lambda_3 x_3.\end{aligned} \tag{25}$$

We may ignore the term $x_1(1-x_2)$ multiplying u, as it is strictly positive for all $x \in \operatorname{int}(C^3)$. Thus, the control system is in the form (22) with $f(x) = [\,\lambda_0(1-x_1) \quad -\lambda_2 x_2 \quad (1-x_3) \quad \lambda_2 x_2(1-x_3) - \lambda_3 x_3\,]'$, and $g(x) = [\,-1 \quad 1 \quad 0\,]'$. A calculation yields

$$v^1 := [f,g] = [\,-\lambda_0 \quad \lambda_2(1-x_3) \quad -\lambda_2(1-x_3)\,]',$$

$$\begin{aligned}v^2 &:= [[f,[f,g]],[f,g]]\\ &= \big[\,0 \quad \lambda_2^2(\lambda_3(2-x_3) - \lambda_2(1-x_3)^2) \quad \lambda_2^2(\lambda_2(1-x_3)^2 - \lambda_3)\,\big]',\end{aligned}$$

$$\begin{aligned}v^3 &:= [v^2, v^1]\\ &= \big[\,0 \quad \lambda_2^3(\lambda_3 x_3 + \lambda_2(1-x_3)^2) \quad -\lambda_2^3(\lambda_2(1-x_3)^2 + \lambda_3)\,\big]',\end{aligned}$$

and

$$\det\big(\big[\,v^1 \quad v^2 \quad v^3\,\big]\big) = 2\lambda_0\lambda_2^5\lambda_3^2(1-x_3).$$

Since this is different from zero for all $x \in \operatorname{int}(C^3)$, we conclude that (25) is accessible from every $x \in \operatorname{int}(C^3)$.

APPENDIX B: PROOFS

Proof of Theorem 1. The proof of (11) follows immediately from Remark 1. Indeed, using the constant control $u(t) \equiv v$ amounts to setting the desired density profile x^f as the steady-state densities of the dynamics, and R^f as the steady-state production rate. Since this steady-state is globally asymptotically stable on $\operatorname{int}(\Omega)$, this implies (11).

We now turn to prove that the system is state- and output-controllable, that is, that we can steer the system to the desired augmented profile $x^f \in \operatorname{int}(C^n), R^f \in \mathbb{R}_{++}$ in finite time. We begin by defining a new control system obtained by replacing $\lambda_i, i \in \{0,\ldots,n-1\}$, in the RFM (3) by a control function $u_i(t) : \mathbb{R}_+ \to \mathbb{R}_+$ (but leaving λ_n as a constant rate). This yields

$$\dot{x} = g_0(x) + \sum_{i=1}^{n} u_{i-1} g_i(x), \tag{26}$$

where $g_0(x) := [\,0 \quad \ldots \quad 0 \quad -\lambda_n x_n\,]'$, $g_1(x) := [\,1-x_1 \quad 0 \ldots 0\,]'$, and for any $j \geq 2$, $g_j(x)$ contains the value $-x_{j-1}(1-x_j)$ in its $(j-1)$'th coordinate, the value $x_{j-1}(1-x_j)$ in its j'th coordinate, and the value 0 otherwise. For example, for $n = 4$

$$\begin{aligned}g_0(x) &= [\,0 \quad 0 \quad 0 \quad -\lambda_4 x_4\,]',\\ g_1(x) &= [\,1-x_1 \quad 0 \quad 0 \quad 0\,]',\\ g_2(x) &= [\,-x_1(1-x_2) \quad x_1(1-x_2) \quad 0 \quad 0\,]',\\ g_3(x) &= [\,0 \quad -x_2(1-x_3) \quad x_2(1-x_3) \quad 0\,]',\\ g_4(x) &= [\,0 \quad 0 \quad -x_3(1-x_4) \quad x_3(1-x_4)\,]'.\end{aligned}$$

Pick $z \in \mathbb{R}^n$. Then it is straightforward to show that

$$z = \sum_{i=1}^{n} \alpha_i g_i(x^f),$$

where

$$\alpha_i := \frac{\sum_{k=i}^{n} z_k}{x_{i-1}^f(1-x_i^f)},$$

with $x_0^f := 1$. Since $x^f \in \operatorname{int}(C^n)$, α_i is well-defined for all $i = 1,\ldots,n$. We conclude that the vector fields $g_1(x^f),\ldots,g_n(x^f)$ span $\mathbb{R}^n$. This implies, by known accessibility results (see, e.g., [58, Ch. 4]), that there exists a set $V = V(x^f) \subseteq \operatorname{int}(C^n)$, that has a nonempty interior in $\mathbb{R}^n$, and such that every $p \in V$ can be steered to x^f in *finite time*. Fix arbitrary $q \in \operatorname{int}(V)$ and $x^s \in C^n$. We already know that there exist constant controls $u_0,\ldots,u_n$ such that $\lim_{t\to\infty} x(t,u,x^s) = q$, $\lim_{t\to\infty} R(t,u,x^s) = R^f$. Therefore there exists a time $\tau > 0$ such that $x(\tau,u,x^s) \in V$. We also know that we can keep u_n at this constant value, and find a time-varying control $w(t) = [\,w_0(t),\ldots, w_{n-1}(t), w_n(t)]$, $t \in [\tau,T]$, with $w_n(t) \equiv u_n$, such that the time-concatenated control steers x^s to x^f at time T. In particular, this control steers $x_n(0) = x_n^s$ to $x_n(T) = x_n^f$. Since u_n is the constant control value such that $R^f = u_n x_n^f$, this yields $R(T) = u_n(T)x_n(T) = R^f$, and this completes the proof. □

Remark 3. Note that the construction above may lead to a production rate $R(t)$ that is discontinuous at $t = 0$. This can be easily overcome using any control $u_n(t)$, $t \in [0,\varepsilon]$, that smoothly interpolates between the value $\frac{R^s}{x_n^s}$ at $t = 0$, and the value $u_n := \frac{R^f}{x_n^f}$ at $t = \varepsilon$. For example, $u_n(t)$ could be picked linear in $t \in [0,\varepsilon]$. We can then apply the constant controls $u_0,\ldots,u_{n-1}$ at $t = \varepsilon$, and continue with the argument above, while noting that now we require $\tau > \varepsilon$.

Proof of Proposition 1. Consider the RFM with rates $\lambda_0,\ldots,\lambda_n$. It was shown in [43, Proposition 1] that R is a strictly increasing function of every λ_i. This means that in order to analyze Ω in the controlled RFM with $u \in \mathbb{U}$ it is enough to consider the reachable set for the controls $u(t) \equiv 0$ and $u(t) \equiv c$. It has been shown in [43] that for the rates $\lambda_0,\ldots,\lambda_n$, the steady state production rate is $R = (\zeta_{MAX}(A(\lambda_0,\ldots,\lambda_n)))^{-2}$. Thus for the two controls above $R(t)$ in the controlled RFM converges to 0 and to $M := (\zeta_{MAX}(A(q_0,\ldots,q_n)))^{-2}$. We conclude that $\Omega(\Theta) = [0,M]$. □

Proof of Proposition 2. Pick $R^f \in \operatorname{int}(\Omega)$. Our goal is to show that there exist a finite time $T \geq 0$ and a control $u \in \mathbb{U}$ that steers $R(t)$ to R^f in time T. We consider two cases.

Case 1. Suppose that $n \notin \Theta$. Since $R^f \in \operatorname{int}(\Omega)$, there exists $\varepsilon > 0$ such that $(R^f - \varepsilon) \in \Omega$ and $(R^f + \varepsilon) \in \Omega$.

Therefore, there exist $v^-, v^+ \in [0, c]$ such that for the control $u^-(t) \equiv v^-$ $[u^+(t) \equiv v^+]$ the production rate converges to $R^f - \varepsilon$ $[R^f + \varepsilon]$ for any x_0. Applying u^- for a sufficiently long time T_1 yields $R(T_1) < R^f$. Now applying u^+ for a sufficiently long time T_2 yields $R(T_1 + T_2) > R^f$. Since $R(t)$ is continuous, this implies that there exists $T \in [T_1, T_1 + T_2]$ such that $R(T) = R^f$.

Case 2. Suppose that $n \in \Theta$, i.e., $R(t) = u(t)x_n(t)$. The argument used in Case 1 does not hold as is because now a discontinuity in u yields a discontinuity in $R(t)$. However, it is clear that we can design a control u by concatenating $u(t) \equiv v^-$ for $t \in [0, T_1]$, then a function of time satisfying $u(T_1) = v^-$ and $u(T_1 + \tau) = v^+$, with $\tau > 0$, and finally $u(t) \equiv v^+$ for $t \geq T_1 + \tau$, and that this will steer $R(t)$ to R^f at some final time T. □

Proof of Proposition 3. It has been shown in [43] that $\frac{\partial R}{\partial \lambda_i}$ exists and is strictly positive for all $i \in \{0, \ldots, n\}$. Combining this with (6) implies that $\frac{\partial e_k}{\partial \lambda_i}$ exists for all $k \in \{1, \ldots, n\}$ and all $i \in \{0, \ldots, n\}$. Pick $i \in \{1, \ldots, n-2\}$. Differentiating (6) with respect to λ_i yields

$$\begin{aligned}
-\lambda_0 e_1' &= \lambda_1 e_1'(1 - e_2) - \lambda_1 e_1 e_2' \\
&= \lambda_2 e_2'(1 - e_3) - \lambda_2 e_2 e_3' \\
&\vdots \\
&= \lambda_{i-1} e_{i-1}'(1 - e_i) - \lambda_{i-1} e_{i-1} e_i' \\
&= e_i(1 - e_{i+1}) + \lambda_i e_i'(1 - e_{i+1}) - \lambda_i e_i e_{i+1}' \\
&= \lambda_{i+1} e_{i+1}'(1 - e_{i+2}) - \lambda_{i+1} e_{i+1} e_{i+2}' \\
&\vdots \\
&= \lambda_{n-1} e_{n-1}'(1 - e_n) - \lambda_{n-1} e_{n-1} e_n' \\
&= \lambda_n e_n' \\
&= R',
\end{aligned} \tag{27}$$

where we use the notation $f' := \frac{\partial f}{\partial \lambda_i}$. Since $R' > 0$, we conclude that $e_1' < 0$. Now the equation $\lambda_1 e_1'(1 - e_2) - \lambda_1 e_1 e_2' = R'$, and the fact that $e \in (0, 1)^n$ yield $e_2' < 0$. Continuing in this fashion yields $e_j' < 0$ for all $j \leq i$. The last equality in (27) yields $\lambda_n e_n' > 0$, so $e_n' > 0$. Now the equality $\lambda_{n-1} e_{n-1}'(1 - e_n) - \lambda_{n-1} e_{n-1} e_n' = R'$ yields $e_{n-1}' > 0$, and continuing in this fashion yields $e_j' > 0$ for all $j > i$. This completes the proof for the case $i \in \{1, \ldots, n-2\}$. The proof when $i \in \{0, n-1, n\}$ is similar. □

ACKNOWLEDGMENTS

We thank Pablo Iglesias for helpful comments. We are grateful to the anonymous reviewers and the AE for comments that helped us to greatly improve this paper. This research is partially supported by research grants from the Israeli Science Foundation, the Israeli Ministry of Science, Technology & Space, the Edmond J. Safra Center for Bioinformatics at Tel Aviv University and the US-Israel Binational Science Foundation. The work of EDS is also supported in part by grants AFOSR FA9550-14-1-0060 and ONR 5710003367. An abridged version of this paper has been presented at the *55th IEEE Conference on Decision and Control* [67].

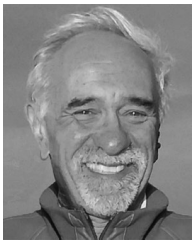
REFERENCES

[1] R. J. R. Algar, T. Ellis, and G. B. Stan, "Modelling essential interactions between synthetic genes and their chassis cell," in *Proc. 53rd IEEE Conf. Decision Control*, 2014, pp. 5437–5444.
[2] A. R. Subramaniam, B. M. Zid, and E. K. O'Shea, "An integrated approach reveals regulatory controls on bacterial translation elongation," *Cell*, vol. 159, no. 5, pp. 1200–1211, 2014.
[3] Y. Arava, Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, and D. Herschlag, "Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae," *Proc. Nat. Academy Sci. United States America*, vol. 100, no. 7, pp. 3889–3894, 2003.
[4] C. Binnie, J. D. Cossar, and D. I. Stewart, "Heterologous biopharmaceutical protein expression in Streptomyces," *Trends Biotechnol.*, vol. 15, no. 8, pp. 315–20, 1997.
[5] R. A. Blythe and M. R. Evans, "Nonequilibrium steady states of matrix-product form: A solver's guide," *J. Physics A: Math. Theoretical*, vol. 40, no. 46, pp. R333–R441, 2007.
[6] F. Brauer, "Compartmental models in epidemiology," in *Mathematical Epidemiology*, F. Brauer, P. van den Driessche, and J. Wu, Eds. Berlin, Germany: Springer, 2008, pp. 19–79.
[7] D. Chu, N. Zabet, and T. von der Haar, "A novel and versatile computational tool to model translation," *Bioinf.*, vol. 28, no. 2, pp. 292–293, 2012.
[8] A. Dana and T. Tuller, "Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells," *PLoS Comput. Biol.*, vol. 8, no. 12, 2012, Art. no. e1002755.
[9] A. Dana and T. Tuller, "Efficient manipulations of synonymous mutations for controlling translation rate–an analytical approach," *J. Comput. Biol.*, vol. 19, pp. 200–231, 2012.
[10] A. Dana and T. Tuller, "Mean of the typical decoding rates: A new translation efficiency index based on ribosome analysis data," *G3: Genes, Genomes, Genetics*, vol. 5, pp. 73–80, 2014.
[11] A. Deana and J. Belasco, "Lost in translation: The influence of ribosomes on bacterial mRNA decay," *Genes Develop.*, vol. 19, no. 21, pp. 2526–33, 2005.
[12] C. Deneke, R. Lipowsky, and A. Valleriani, "Effect of ribosome shielding on mRNA stability," *Phys. Biol.*, vol. 10, no. 4, 2013, Art. no. 046008.
[13] A. Diament and T. Tuller, "Estimation of ribosome profiling performance and reproducibility at various levels of resolution," *Biol. Direct*, vol. 11, 2016, Art. no. 24.
[14] M. Dreyfus and S. A. Joyce, "The interplay between translation and mRNA decay in prokaryotes: A discussion on current paradigms," in *Translation Mechanisms*, J. Lapointe and L. Brakier-Gingras, Eds. Norwell, MA, USA: Kluwer, 2002, pp. 165–183.
[15] D. A. Drummond and C. O. Wilke, "Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution," *Cell*, vol. 134, pp. 341–352, 2008.
[16] S. Edri, E. Gazit, E. Cohen, and T. Tuller, "The RNA polymerase flow model of gene transcription," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 1, pp. 54–64, Feb. 2014.
[17] S. Edri and T. Tuller, "Quantifying the effect of ribosomal density on mRNA stability," *PLoS One*, vol. 9, 2014, Art. no. e102308.
[18] B. Gorgoni, E. Marshall, M. McFarland, M. Carmen Romano , and I. Stansfield, "Controlling translation elongation efficiency: tRNA regulation of ribosome flux on the mRNA," *Biochemical Soc. Trans.*, vol. 42, no. 1, pp. 160–165, 2014.
[19] Y. Hayakawa, S. Hosoe, M. Hayashi, and M. Ito, "On the structural controllability of compartmental systems," *IEEE Trans. Autom. Control*, vol. 29, no. 1, pp. 17–24, Jan. 1984.
[20] M. Holza and A. Fahrb, "Compartment modeling," *Adv. Drug Del. Rev.*, vol. 48, pp. 249–264, 2001.
[21] A. C. Hsieh, et al., "The translational landscape of mTOR signalling steers cancer initiation and metastasis," *Nature*, vol. 485, pp. 55–61, 2012.
[22] R. Jackson, C. Hellen, and T. Pestova, "The mechanism of eukaryotic translation initiation and principles of its regulation," *Nature Rev. Mol. Cell Biol.*, vol. 11, no. 2, pp. 113–27, 2010.
[23] J. A. Jacquez, *Compartmental Analysis in Biology and Medicine*, 3rd ed. Ann Arbor, MI, USA: BioMedware, 1996.
[24] J. A. Jacquez and C. P. Simon, "Qualitative theory of compartmental systems," *SIAM Rev.*, vol. 35, no. 1, pp. 43–79, 1993.
[25] L. É. Johnson, "Control and controllability in compartmental systems," *Math. Biosci.*, vol. 30, no. 1/2, pp. 181–190, 1976.
[26] M. Kaern, T. Elston, W. Blake, and J. Collins, "Stochasticity in gene expression: From theories to phenotypes," *Nature Rev. Genetics*, vol. 6, no. 6, pp. 451–64, 2005.

[27] C. Kimchi-Sarfaty, et al., "A "silent" polymorphism in the MDR1 gene changes substrate specificity," *Sci.*, vol. 315, pp. 525–528, 2007.
[28] C. Kimchi-Sarfaty, T. Schiller, N. Hamasaki-Katagiri, M. A. Khan, C. Yanover, and Z. E. Sauna, "Building better drugs: Developing and regulating engineered therapeutic proteins," *Trends Pharmacological Sci.*, vol. 34, no. 10, pp. 534–548, 2013.
[29] C. Kurland, "Translational accuracy and the fitness of bacteria," *Annu. Rev. Genetics*, vol. 26, pp. 29–50, 1992.
[30] C.-T. Lin, "Structural controllability," *IEEE Trans. Automat. Control*, vol. 19, no. 3, pp. 201–208, Jun. 1974.
[31] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabasi, "Controllability of complex networks," *Nature*, vol. 473, pp. 167–173, 2011.
[32] F. Loayza-Puch, J. Drost, K. Rooijers, R. Lopes, R. Elkon, and R. Agami, "p53 induces transcriptional and translational programs to suppress cell proliferation and growth," *Genome Biol.*, vol. 14, no. 4, pp. 1–12, 2013.
[33] M. Margaliot, E. D. Sontag, and T. Tuller, "Entrainment to periodic initiation and transition rates in a computational model for gene translation," *PLOS One*, vol. 9, no. 5, 2014, Art. no. e96039.
[34] M. Margaliot and T. Tuller, "Stability analysis of the ribosome flow model," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 5, pp. 1545–1552, Sep./Oct. 2012.
[35] M. Margaliot and T. Tuller, "Ribosome flow model with positive feedback," *J. Roy. Soc. Interface*, vol. 10, 2013, Art. no. 20130267.
[36] H. Mayeda and T. Yamada, "Strong structural controllability," *SIAM J. Control Optimization*, vol. 17, no. 1, pp. 123–138, 1979.
[37] T. Moks, et al., "Expression of human insulin-like growth factor I in bacteria: Use of optimized gene fusion vectors to facilitate protein purification," *Biochemistry*, vol. 26, no. 17, pp. 5239–44, 1987.
[38] D. A. Nissley, et al., "Accurate prediction of cellular co-translational folding indicates proteins can switch from post-to co-translational folding," *Nature Commun.*, vol. 7, 2016, Art. no. 10341.
[39] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 3, pp. 249–258, Sep. 2014.
[40] S. Pechmann and J. Frydman, "Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding," *Nature Structural Mol. Biol.*, vol. 20, no. 2, pp. 237–43, 2013.
[41] M. Pedersen, S. Nissen, N. Mitarai, S. Lo Svenningsen, K. Sneppen, and S. Pedersen, "The functional half-life of an mRNA depends on the ribosome spacing in an early coding region," *J. Mol. Biol.*, vol. 407, no. 1, pp. 35–44, 2011.
[42] J. B. Plotkin and G. Kudla, "Synonymous but not the same: The causes and consequences of codon bias," *Nature Rev. Genetics*, vol. 12, pp. 32–42, 2010.
[43] G. Poker, Y. Zarai, M. Margaliot, and T. Tuller, "Maximizing protein translation rate in the nonhomogeneous ribosome flow model: A convex optimization approach," *J. Roy. Soc. Interface*, vol. 11, no. 100, 2014, Art. no. 20140713.
[44] G. Poker, M. Margaliot, and T. Tuller, "Sensitivity of mRNA translation," *Sci. Rep.*, vol. 5, 2015, Art. no. 12795.
[45] S. Proshkin, A. Rahmouni, A. Mironov, and E. Nudler, "Cooperation between translating ribosomes and RNA polymerase in transcription elongation," *Sci.*, vol. 328, no. 5977, pp. 504–508, 2010.
[46] J. Racle, F. Picard, L. Girbal, M. Cocaign-Bousquet, and V. Hatzimanikatis, "A genome-scale integration and analysis of Lactococcus lactis translation data," *PLOS Comput. Biol.*, vol. 9, no. 10, 2013, Art. no. e1003240.
[47] A. Raveh, Y. Zarai, M. Margaliot, and T. Tuller, "Ribosome flow model on a ring," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 6, pp. 1429–1439, Nov./Dec. 2015.
[48] A. Raveh, M. Margaliot, E. D. Sontag, and T. Tuller, "A model for competition for ribosomes in the cell," *J. Roy. Soc. Interface*, vol. 13, 2016, Art. no. 20151062.
[49] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppin, and T. Tuller, "Genome-scale analysis of translation elongation with a ribosome flow model," *PLOS Comput. Biol.*, vol. 7, 2011, Art. no. e1002127.
[50] M. A. Romanos, C. A. Scorer, and J. J. Clare, "Foreign gene expression in yeast: A review," *Yeast*, vol. 8, no. 6, pp. 423–488, 1992.
[51] R. Sabi and T. Tuller, "A comparative genomics study on the effect of individual amino acids on ribosome stalling," *BMC Genomics*, vol. 16, 2015, Art. no. S5.
[52] H. Salis, E. Mirsky, and C. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnol.*, vol. 27, no. 10, pp. 946–50, 2009.
[53] A. Schadschneider, D. Chowdhury, and K. Nishinari, *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Amsterdam, The Netherlands: Elsevier, 2011.
[54] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, and J. Plotkin, "Rate-limiting steps in yeast protein translation," *Cell*, vol. 153, no. 7, pp. 1589–601, 2013.
[55] L. B. Shaw, R. K. P. Zia, and K. H. Lee, "Totally asymmetric exclusion process with extended objects: A model for protein synthesis," *Phys. Rev. E*, vol. 68, 2003, Art. no. 021910.
[56] D. Siljak, *Large Scale Dynamic Systems*. New York, NY, USA: North-Holland, 1978.
[57] H. L. Smith, *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. Providence, RI, USA: Amer. Math. Soc., 1995, vol. 41.
[58] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd ed. New York, NY, USA: Springer, 1998.
[59] L. Sun, et al., "A recombinant collagen-mRNA platform for controllable protein synthesis," *ChemBioChem*, vol. 16, no. 10, pp. 1415–9, 2015.
[60] V. A. Tomlinson, et al., "Translation elongation factor eEF1A2 is a potential oncoprotein that is overexpressed in two-thirds of breast tumours," *BMC Cancer*, vol. 5, no. 1, pp. 1–7, 2005.
[61] T. Tuller, et al., "An evolutionarily conserved mechanism for controlling the efficiency of protein translation," *Cell*, vol. 141, no. 2, pp. 344–54, 2010.
[62] T. Tuller, I. Veksler, N. Gazit, M. Kupiec, E. Ruppin, and M. Ziv, "Composite effects of the coding sequences determinants on the speed and density of ribosomes," *Genome Biol.*, vol. 12, no. 11, 2011, Art. no. R110.
[63] T. Tuller and H. Zur, "Multiple roles of the coding sequence 5′ end in gene expression regulation," *Nucleic Acids Res.*, vol. 43, no. 1, pp. 13–28, 2015.
[64] T. Tuller, M. Kupiec, and E. Ruppin, "Determinants of protein abundance and translation efficiency in S. cerevisiae," *PLOS Comput. Biol.*, vol. 3, pp. 1–10, 2007.
[65] Y. Y. Waldman, T. Tuller, R. Sharan, and E. Ruppin, "TP53 cancerous mutations exhibit selection for translation efficiency," *Cancer Res.*, vol. 69, pp. 8807–8813, 2009.
[66] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, New York, NY, USA: Garland Science, 2002.
[67] Y. Zarai, M. Margaliot, E. D. Sontag, and T. Tuller, "Controlling the ribosomal density profile in mRNA translation," in *Proc. 55th IEEE Conf. Decision Control*, 2016, pp. 4184–4189.
[68] Y. Zarai, M. Margaliot, and T. Tuller, "Explicit expression for the steady-state translation rate in the infinite-dimensional homogeneous ribosome flow model," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, no. 5, pp. 1322–1328, Sep./Oct. 2013.
[69] Y. Zarai, M. Margaliot, and T. Tuller, "On the ribosomal density that maximizes protein translation rate," *PLOS One*, vol. 11, no. 11, pp. 1–26, Nov. 2016.
[70] G. Zhang, M. Hubalewska, and Z. Ignatova, "Transient ribosomal attenuation coordinates protein synthesis and co-translational folding," *Nature Structural Mol. Biol.*, vol. 16, no. 3, pp. 274–80, 2009.
[71] R. Zia, J. Dong, and B. Schmittmann, "Modeling translation in protein synthesis with TASEP: A tutorial and recent developments," *J. Statistical Physics*, vol. 144, pp. 405–428, 2011.
[72] H. Zur and T. Tuller, "New universal rules of eukaryotic translation initiation fidelity," *PLOS Comput. Biol.*, vol. 9, no. 7, 2013, Art. no. e1003136.
[73] H. Zur and T. Tuller, "Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution," *Nucleic Acids Res.*, vol. 44, no. 19, pp. 9031–9049, 2016.

Yoram Zarai received the BSc (cum laude), MSc, and PhD degrees in electrical engineering from Tel Aviv University, in 1992, 1998, and 2017, respectively. He is currently a postdoctoral researcher with Tel Aviv University. His research interests include modeling and analysis of biological phenomena, machine learning, and signal processing.

Michael Margaliot received the BSc (cum laude) and MSc degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1992 and 1995, respectively, and the PhD (summa cum laude) degree from Tel Aviv University, in 1999. He was a post-doctoral fellow in the Department of Theoretical Mathematics, Weizmann Institute of Science. In 2000, he joined the Department of Electrical Engineering-Systems, Tel Aviv University, where he is currently a professor and chair. His research interests include the stability analysis of differential inclusions and switched systems, optimal control theory, fuzzy control, computation with words, Boolean control networks, and systems biology. He is co-author of *New Approaches to Fuzzy Modeling and Control: Design and Analysis*, World Scientific, 2000 and of *Knowledge-Based Neurocomputing: A Fuzzy Logic Approach*, Springer, 2009. He currently serves as associate editor of the journal *IEEE Transactions on Automatic Control*.



Eduardo D. Sontag received the undergraduate degree in mathematics from the University of Buenos Aires, in 1972 and the PhD degree in mathematics from the University of Florida, in 1976, working under Rudolf E. Kalman. Since 1977, he has been with Rutgers University, where he is a distinguished professor in the Department of Mathematics. He is also a member of the Rutgers Cancer Institute of New Jersey as well as of the graduate faculties of the Departments of Computer Science and of Electrical and Computer Engineering. He currently serves as head of the Undergraduate Biomathematics Interdisciplinary Major, director of the Graduate Program in Quantitative Biomedicine, and Director of the Center for Quantitative Biology. His current research interests include broadly in applied mathematics, and specifically in systems biology, dynamical systems, and feedback control theory. In the 1980s and 1990s, he introduced new tools for analyzing the effect of external inputs on the stability of nonlinear systems ("input to state stability") and for feedback design ("control-Lyapunov functions"), both of which have been widely adopted as paradigms in engineering research and education. He also developed the early theory of hybrid (discrete/continuous) control, and worked on learning theory applied to neural processing systems as well as in foundations of analog computing. Starting around 1999, his work has turned in large part to developing basic theoretical aspects of biological signal transduction pathways and gene networks, as well as collaborations with a range of experimental and computational biological labs dealing with cell cycle modeling, development, cancer progression, infectious diseas es, physiology, synthetic biology, and other topics. He has published about 500 papers in fields ranging from control theory and theoretical computer science to cell biology, with more than 33,000 citations and a (google scholar) h-index of 81. He was awarded the Reid Prize by SIAM in 2001, the Bode Prize in 2002 and the Control Systems Field Award in 2011 from IEEE, and the 2002 Board of Trustees Award for Excellence in Research and the 2005 Teacher/Scholar Award from Rutgers. He is a fellow of the IEEE, the AMS, the SIAM, and the IFAC.

Tamir Tuller received the BSc degree in electrical engineering, mechanical engineering and computer science from Tel Aviv University, Tel Aviv, Israel, the MSc degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, and two PhD degrees, one in computer science and one in medical science, from Tel Aviv University. He was a Safra postdoctoral fellow in the School of Computer Science and the Department of Molecular Microbiology and Biotechnology, Tel Aviv University, and a Koshland postdoctoral fellow in the Faculty of Mathematics and Computer Science, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel. In 2011, he joined the Department of Biomedical Engineering, Tel Aviv University, where he is currently an associate professor. His research interests fall in the general areas of synthetic biology, systems biology, and computational biology. In particular, he works on deciphering, modeling, and engineering of gene expression.