Computational Tool

EGGTART: A tool to visualize the dynamics of biophysical transport under the inhomogeneous l-TASEP

Dan D. Erdmann-Pham,[1] Wonjun Son,[2] Khanh Dao Duc,[3,*] and Yun S. Song[4,5,6,*]

[1]Department of Mathematics, University of California, Berkeley, California; [2]Department of Computer Science, Columbia University, New York, New York; [3]Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada; [4]Computer Science Division and [5]Department of Statistics, University of California, Berkeley, California; and [6]Chan Zuckerberg Biohub, San Francisco, California

ABSTRACT   The totally asymmetric simple exclusion process (TASEP), which describes the stochastic dynamics of interacting particles on a lattice, has been actively studied over the past several decades and applied to model important biological transport processes. Here, we present a software package, called EGGTART (Extensive GUI gives TASEP-realization in Real Time), which quantifies and visualizes the dynamics associated with a generalized version of the TASEP with an extended particle size and heterogeneous jump rates. This computational tool is based on analytic formulas obtained from deriving and solving the hydrodynamic limit of the process. It allows an immediate quantification of the particle density, flux, and phase diagram, as a function of a few key parameters associated with the system, which would be difficult to achieve via conventional stochastic simulations. Our software should therefore be of interest to biophysicists studying general transport processes and can in particular be used in the context of gene expression to model and quantify mRNA translation of different coding sequences.

SIGNIFICANCE   Stochastic transport phenomena on inhomogeneous lattices are complex and typically accessible only through simulations or approximations, as closed-form formulas are often intractable. Recently, we obtained the stationary solution to the so-called inhomogeneous ℓ-TASEP, a central model of biophysical transport, in the limit of large lattice length. We found that ultimately, there are only six parameters contributing substantially to particle currents and densities—the key quantities associated with transport efficiency. Based on this theoretical advance, we present here a computational tool to visualize and interactively explore the relationship between the key system parameters we have identified and steady-state densities and currents, thus enabling one to explore the robustness and optimality of transport systems.

INTRODUCTION

The totally asymmetric exclusion process (TASEP) is a stochastic process used to model a large variety of transport phenomena involving interacting particles (1). Although it has been studied extensively over the past several decades, it is still the subject of active research and is replete with many open problems motivated by biophysical applications. In particular, the recent emergence of experimental data through advances in microscopy and sequencing techniques has re-

vealed a wealth of molecular processes that are well described by TASEP models, including the motion of ribosomes (2–4), RNA polymerase (5), and motor proteins (6). Such processes require extending the classical TASEP by several layers of complexity, and as a result, their analysis has remained elusive and restricted to particular cases that do not adequately reflect biology. This shortcoming recently led us to study the so-called inhomogeneous ℓ-TASEP, in which particles of size ℓ traverse a lattice of inhomogeneous jump rates. By considering the hydrodynamic limit of the process, we obtained exact formulas for particle currents and locally averaged occupation probabilities (7), offering immediate quantification of the particle density, flux, and phase diagram, which would be difficult to achieve via conventional stochastic simulations.

To help to visualize these theoretical results, we present here a software package called EGGTART (extensive GUI

https://doi.org/10.1016/j.bpj.2021.02.004

gives TASEP realization in real time). For an arbitrary specification of the key system parameters (particle size, entrance rate, exit rate, and site-specific jump rates), EGGTART provides a graphical user interface that allows us to extract the main quantities of interest (e.g., the current and local densities, i.e., smoothed occupation probabilities, of particles), in addition to providing a phase diagram that fully describes different types of traffic behavior. We demonstrate how it can recover various versions of the TASEP studied in the past literature as special cases. By facilitating the exploration and visualization of both theoretical and practical aspects of the inhomogeneous ℓ-TASEP, we anticipate our computational tool to be broadly helpful to mathematicians, physicists, and biologists in studying various biophysical transport phenomena.

METHODS

The inhomogeneous ℓ-TASEP

The inhomogeneous ℓ-TASEP is a Markov process, illustrated in Fig. 1, in which particles of size ℓ jump unidirectionally along a lattice of N sites under mutual exclusion: a particle at site $i \in \{1, ..., N-1\}$ remains at its position if the site $i + \ell$ is occupied and jumps at exponential rate p_i to position $i + 1$ otherwise. Particles can enter the lattice at the first site at rate α while respecting the mutual exclusion constraint and leave the lattice at the last site at rate β.

The hydrodynamic limit

Although solving the inhomogeneous ℓ-TASEP is in general intractable, we can do so in the continuum limit of $N \to \infty$. More precisely, denoting by $\tau(t) \in \{0, 1\}^N$ a configuration at time t and assuming that $\lambda(x) = \lim_{N \to \infty} \frac{1}{\ell} \sum_{i=0}^{\ell-1} p_{\lfloor Nx \rfloor + i}$ exists and is differentiable, we have shown in (7) that the ℓ-smoothed density of particles $\rho(x, t) dx = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N-\ell+1} \frac{1}{\ell} \sum_{i=0}^{\ell-1} \tau_{n+i}(Nt) \delta_{\frac{n+i}{N}}(dx)$ satisfies the inhomogeneous conservation law

$$\partial_t \rho = -\partial_x [\lambda(x) \rho \, G(\rho)], \qquad (1)$$

where $G(\rho) = (1 - \ell\rho)/(1 - (\ell - 1)\rho)$. Constant jump rates and $\ell = 1$ yield the one-dimensional Burgers' equation, known as the hydrodynamic limit of the classical TASEP (8). As shown next, increasing ℓ and introducing spatial inhomogeneity lead to more complex solutions.

Analytical solutions and phase diagram

Equation 1 can be solved using the method of characteristics (7), wherein the two boundary points 0 and 1 each emit a characteristic curve whose evolution determines the long-term behavior of the system. These curves are controlled by boundary conditions, which yield a phase diagram in α and β. Surprisingly, this phase diagram is completely characterized by four parameters: the particle size ℓ and the minimal, initial, and terminal jump rates $\lambda_{\min} := \min_x \lambda(x)$, $\lambda_0 := \lambda(0)$, and $\lambda_1 := \lambda(1)$, respectively. These parameters determine the critical rates $\alpha^* = \alpha^*(\ell, \lambda_0, \lambda_{\min})$ and $\beta^* = \beta^*(\ell, \lambda_1, \lambda_{\min})$ at which phase transitions occur (precise expressions are provided in (7)). The different regions of the phase diagram are characterized as follows (note
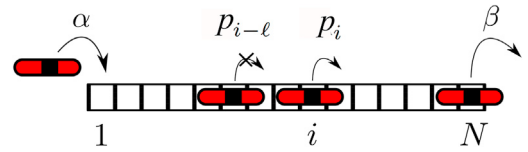

FIGURE 1 Illustration of the inhomogeneous ℓ-TASEP with open boundaries. Particles (of size $\ell = 3$ here) enter the first site of the lattice at rate α, and a particle at position i (here denoted by the position of the midpoint of the particle) moves one site to the right at rate p_i, provided that site $i + \ell$ is empty. A particle at the end of the lattice exits at rate β. To see this figure in color, go online.

that Eq. 1 cannot describe critical behavior, in which residual stochasticity persists; see (7) and references therein):

1) $\alpha < \alpha^*$ and $\beta > \beta^*$: In this low-density regime (LD), the low entrance rate coupled with large exit rate establish an overall small density $\rho_L(x) = \rho_L(\lambda(x), \ell, J_L) < \rho^* := (\ell + \sqrt{\ell})^{-1}$, where $J_L = \alpha(\lambda_0 - \alpha)/(\lambda_0 + (\ell - 1)\alpha)$ is the particle current.

2) $\alpha > \alpha^*$ and $\beta < \beta^*$: Particles are injected frequently and drained slowly, shaping the high-density phase (HD). The associated density $\rho_R(x) = \rho_R(\lambda(x), \ell, J_R) > \rho^*$ is strictly larger than ρ_L, with a current now given by $J_R = \beta(\lambda_1 - \beta)/(\lambda_1 + (\ell - 1)\beta)$.

3) $\alpha < \alpha^*$ and $\beta < \beta^*$: A phase transition occurs along the nonlinear curve $J_L = J_R$, with $J_L < J_R$ resulting in LD densities and currents and $J_L > J_R$ in HD ones.

4) $\alpha > \alpha^*$ and $\beta > \beta^*$: This phase characterizes the maximal current regime (MC). The maximal current $J_{\max} = \lambda_{\min} \times (1 + \sqrt{\ell})^{-2}$ results from of superposition of high and low densities $\rho(x) = \rho_R(x) \times \mathbb{1}_{x \leq x_{\min}} + \rho_L(x) \times \mathbb{1}_{x \geq x_{\min}}$, where $x_{\min} = \mathrm{argmin}_x \lambda(x)$.

General description of the software and availability

For any parameterization of the inhomogeneous ℓ-TASEP (e.g., particle size, entrance, exit, and site-specific rates), our computational tool EGGTART integrates the analytical formulas given above into a graphical user interface that provides a quantification and visualization of key properties, such as the particle current, densities, and phase diagram associated with different types of particle traffic behavior. It has been developed in Python 3.6 and uses the numpy and pyqtgraph packages. Versions for Mac OS X, Linux (Ubuntu), and Windows (10) are available at https://github.com/songlab-cal/eggtart, together with a user manual describing general features and a tutorial to be used with the demo .csv input files.

RESULTS

We demonstrate the utility of EGGTART for a wide range of applications and examples, ranging in order of increasing complexity from the original TASEP model to the general inhomogeneous ℓ-TASEP. Illustrations of EGGTART's accuracy through Monte Carlo simulations are provided in the Supporting materials and methods.

The homogeneous ℓ-TASEP

The classical TASEP model, for homogeneous rates ($p_i \equiv 1$) and $\ell = 1$, has been analytically solved through the matrix ansatz method (9). In the top left panel of
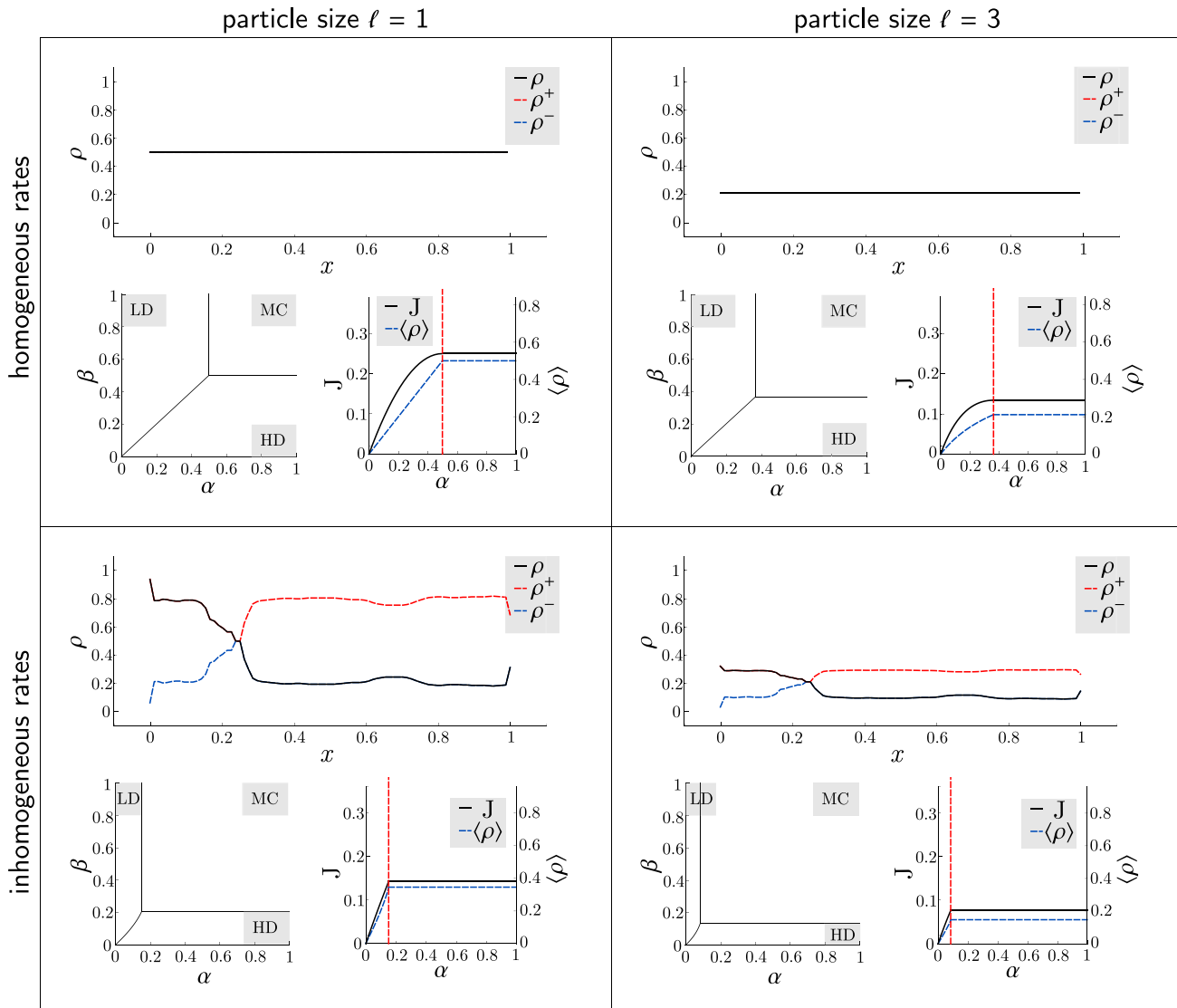
FIGURE 2 The effect of inhomogeneous jump rates and extended particles. The various impacts of increasing particle sizes and introducing inhomogeneity on the phase diagram, density profile, and current are displayed using EGGTART. Particle size and rates can be adjusted independently to investigate individual (*top right* and *bottom left*) and joint effects (*bottom right*). To see this figure in color, go online.

Fig. 2, we show that EGGTART recovers the classical results obtained therein as special cases of the general model; because of the symmetry of particles and holes, the resulting phase diagram is symmetric in α and β, with three regions (LD, HD, and MC; see Methods) separated by simple lines intersecting at the critical point $(\alpha^*, \beta^*) = (1/2, 1/2)$. Particle currents J in LD, HD and MC are $\alpha(1 - \alpha)$, $\beta(1 - \beta)$, and $1/4$, respectively, with bulk densities ρ given by α, $1 - \beta$, and $1/2$.

General transport phenomena typically require particle size $\ell > 1$; e.g., ribosomes translating mRNA sequences occupy roughly 10 codons. The ℓ-TASEP for arbitrary $\ell \geq 1$ models this effect. As in the case for $\ell = 1$, mean-field approaches can provide approximations to the phase diagram, currents, and densities (10–12), which are made precise in the hydrodynamic limit (13). EGGTART also recovers these results, as seen in the top right panel of Fig. 2. The critical values α^* and β^* shrink to $1/(1 + \sqrt{\ell})$, whereas J decreases to $\alpha(1 - \alpha)/(1 + (\ell - 1)\alpha)$, $\beta(1 - \beta)/(1 + (\ell - 1)\beta)$, and $1/(1 + \sqrt{\ell})^2$ in LD, HD, and MC, respectively (with similar adjustment for ρ), showing that increasing the size of particles leads to a global decrease of densities, currents, and transport capacity.

Intermediate cases: bottlenecks and linear rate function

Simple lattice inhomogeneities have been studied extensively, as they promise closer modeling of actual transport phenomena while maintaining theoretical tractability. The

most fundamental examples include single defects or defect clusters, i.e., sites with lower rates, that can model "slow" codons in translation or structural imperfections of the structure in microtubular transport (14,15). Processes with monotonically varying jump rates, which are most simply modeled by linear functions, constitute another class of examples of inhomogeneity and have notably been observed in translation as the so-called "5′ translational ramp" (16). EGGTART readily visualizes and quantifies such configurations, reproducing previously reported results. We provide more details on these intermediate examples and their corresponding biophysical applications in the Supporting materials and methods.

The inhomogeneous ℓ-TASEP

The most general case treated by our model combines extended particles and heterogeneous rates. Such a generalization is, for example, necessary to properly model mRNA translation because the local elongation rate of ribosomes (which occupy 10 codons) depends on multiple factors (17). Therefore, most realistic modeling of translation dynamics has been based on the inhomogeneous ℓ-TASEP (2), with estimates of rates obtained using measurements of transfer RNA usage (2,18) or, more recently, by analyzing ribosome profiling data (17). Until recently (7,19), detailed analyses under such heterogeneous lattices have been restricted to numerical simulations (2) or mean-field approximations suffering from numerical instabilities and imprecisions (20). With EGGTART, we are able to quantify and discern from each other effects due to particle size and inhomogeneity; we observe 1) the reduction in transport capacity and critical rates (α^* and β^*) associated with a decrease of the limiting jump rate $\lambda_{\min}$ (*bottom left panel* in Fig. 2) and extended particles (*bottom right panel*), 2) deformation of the LD-HD phase separation (*bottom panels*), 3) the branch switching of ρ in MC (*bottom panels*), and 4) the impact of λ_0 on the sensitivity of particle current to the initiation rate (see the Supporting materials and methods). Each of these principles has important consequences for optimizing ribosome usage and translation efficiency (7). Therefore, using the software can help practitioners understand a gene's evolutionary constraints or devise sequences most suitable for specific needs.

DISCUSSION

Although the TASEP has been widely used and studied for several decades (1,2,8), we provide here, to the best of our knowledge, the first computational tool for visualizing phase transitions, densities, and fluxes of the TASEP in full generality, with an arbitrary particle size and inhomogeneous jump rates. Our tool has several advantages compared to numerical simulations. By tuning the parameters of the model through a graphical interface, the user can quickly visualize their impact, particularly that of the six key parameters that govern transport efficiency (7). In the context of translation, exploring these parameters allows to address various biological questions, like quantifying how far individual genes are from their transport capacity, and to which parameters protein production is most sensitive (7). As a computational tool that allows immediate quantification and fine-tuning of the TASEP, EGGTART can thus aid in studying the properties of biophysical transport systems while serving as a basis for future improvements and extensions (e.g., the inclusion of Langmuir kinetics (21)).

AUTHOR CONTRIBUTIONS

Y.S.S., K.D.D., and D.D.E.-P. designed the research. W.S. wrote the software. Y.S.S., K.D.D., and D.D.E.-P. wrote the article.

ACKNOWLEDGMENTS

REFERENCES

1. Schadschneider, A., D. Chowdhury, and K. Nishinari. 2010. Stochastic Transport in Complex Systems: from Molecules to Vehicles. Elsevier, Amsterdam, the Netherlands.

2. Zur, H., and T. Tuller. 2016. Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. *Nucleic Acids Res.* 44:9031–9049.

3. Korkmazhan, E., H. Teimouri, …, E. Levine. 2017. Dynamics of translation can determine the spatial organization of membrane-bound proteins and their mRNA. *Proc. Natl. Acad. Sci. USA.* 114:13424–13429.

4. Szavits-Nossan, J., L. Ciandrini, and M. C. Romano. 2018. Deciphering mRNA sequence determinants of protein production rate. *Phys. Rev. Lett.* 120:128101.

5. van den Berg, A. A., and M. Depken. 2017. Crowding-induced transcriptional bursts dictate polymerase and nucleosome density profiles along genes. *Nucleic Acids Res.* 45:7623–7632.

6. Miedema, D. M., V. S. Kushwaha, …, P. Schall. 2017. Correlation imaging reveals specific crowding dynamics of kinesin motor proteins. *Phys. Rev. X.* 7:041037.

7. Erdmann-Pham, D. D., K. Dao Duc, and Y. S. Song. 2020. The key parameters that govern translation efficiency. *Cell Syst.* 10:183–192.e6.

8. Blythe, R. A., and M. R. Evans. 2007. Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A Math. Theor.* 40:R333–R441.

9. Derrida, B., M. R. Evans, …, V. Pasquier. 1993. Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J. Phys. A Math. Gen.* 26:1493–1517.

10. Lakatos, G., and T. Chou. 2003. Totally asymmetric exclusion processes with particles of arbitrary size. *J. Phys. A Math. Gen.* 36:2027–2041.

11. Shaw, L. B., R. K. Zia, and K. H. Lee. 2003. Totally asymmetric exclusion process with extended objects: a model for protein synthesis. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:021910.

12. Dao Duc, K., Z. H. Saleem, and Y. S. Song. 2018. Theoretical analysis of the distribution of isolated particles in totally asymmetric exclusion processes: application to mRNA translation rate estimation. *Phys. Rev. E.* 97:012106.

13. Schönherr, G. 2005. Hard rod gas with long-range interactions: exact predictions for hydrodynamic properties of continuum systems from discrete models. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71:026122.

14. Pierobon, P., M. Mobilia, …, E. Frey. 2006. Bottleneck-induced transitions in a minimal model for intracellular transport. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 74:031906.

15. Dong, J. J., B. Schmittmann, and R. K. Zia. 2007. Inhomogeneous exclusion processes with extended objects: the effect of defect locations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76:051113.

16. Tuller, T., A. Carmi, …, Y. Pilpel. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell.* 141:344–354.

17. Dao Duc, K., and Y. S. Song. 2018. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLoS Genet.* 14:e1007166.

18. Dana, A., and T. Tuller. 2012. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.* 8:e1002755.

19. Szavits-Nossan, J., M. C. Romano, and L. Ciandrini. 2018. Power series solution of the inhomogeneous exclusion process. *Phys. Rev. E.* 97:052139.

20. Shaw, L. B., J. P. Sethna, and K. H. Lee. 2004. Mean-field approaches to the totally asymmetric exclusion process with quenched disorder and large particles. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70:021901.

21. Kuan, H.-S., and M. D. Betterton. 2016. Phase-plane analysis of the totally asymmetric simple exclusion process with binding kinetics and switching between antiparallel lanes. *Phys. Rev. E.* 94:022419.