Enhanced Sampling in Spatial Stochastic Systems Biology Models Using a Weighted Ensemble of Trajectories

Rory M. Donovan[1], Jose Juan Tapia[1], Devin P. Sullivan[1], James R. Faeder[2], Robert F. Murphy[3], Markus T. Dittrich[4], and Daniel M. Zuckerman[2]

***Short Abstract* — We demonstrate that the weighted ensemble (WE) sampling strategy, initially developed for molecular simulations, can be effectively employed for spatial cell-scale simulations. Here, we use WE to orchestrate kinetic Monte Carlo simulations of the MCell platform, which include spatial geometry (e.g. organelles, membranes) and biochemical interactions among mobile molecular species. We study a series of models exhibiting spatial, temporal, and biochemical complexity and show that WE can achieve performance significantly exceeding standard parallel simulation by orders of magnitude for measuring certain observables.**

***Keywords* — Stochastic Dynamics, Monte Carlo, Signaling Networks, Enhanced Sampling, Weighted Ensemble.**

I. Introduction

STOCHASTIC effects are of crucial importance in many biological processes, from protein dynamics, to gene expression, to population-level phenotypic heterogeneity [1,2]. Unfortunately, due to the high computational cost of simulating complex stochastic systems, the effects of stochasticity on system response remain under-studied in complex, realistic biological models.

Spatial models of stochastic reaction-diffusion processes have found widespread use as tools in understanding the mechanics of biological processes on the cellular level and beyond [3]. Regrettably, exhaustively simulating large, realistic models and extracting well-sampled values of experimentally relevant quantities is often beyond the current realm of computational feasibility.

Enhanced sampling algorithms offer an attractive resolution to the dilemma of sampling complex systems: instead of compromising on model complexity in order to achieve well sampled results, rather use one's simulation resources more effectively and extract more information given the same resources. There has been significant interest in sampling algorithms in the field of protein simulation; arguably, such approaches have transformed the field of molecular simulation [4].

[1]Joint CMU-Pitt Ph.D. Program in Computational Biology
[2]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh
[3]Department of Computational Biology, School of Computer Science, Carnegie Mellon University
[4]Pittsburgh Supercomputing Center
E-mail: donovanr@pitt.edu

II. Results

Here, we demonstrate a method to drastically decrease the cost of simulating spatial models of stochastic cellular systems, by applying the weighted ensemble sampling procedure [5]. The WE approach runs an ensemble of parallel trajectories and uses a statistical strategy of replicating and pruning trajectories to focus computational effort on difficult-to-sample regions, which it uses to generate unbiased estimates of non-equilibrium and equilibrium observables.

We present initial results for a toy diffusive binding system, as well as two more complex systems: a cross-compartmental signal transduction model in a realistic cellular geometry, and a model of an active zone in a frog neuromuscular junction. We demonstrate speedups of many orders of magnitude in sampling these models of cellular behavior with spatial dependence.

III. Conclusion

We are able to sample the rare events and full probability distributions for stochastic systems biology models over a wide range of complexity. We demonstrate speed-ups over brute-force sampling that are dramatic enough to encourage the design of more complex, more realistic models. Long time-scale behavior can be extrapolated from short simulations, providing a bridge between dynamics over multiple time-scales. Weighted ensemble is an ideal approach to employ in addressing the issue of difficult-to-sample stochastic systems, and we anticipate further applications to more realistic systems.

References

[1] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. 94, 814 (1997)
[2] B. Munsky, G. Neuert, and A. van Oudenaarden, Science 336, 6078 (2012).
[3] R. A. Kerr, et al. SIAM Journal on Scientific Computing 30.6 3126-3149 (2008).
[4] M. C. Zwier and L. T. Chong, Curr. Opin. Pharmacol. 10, 745 (2010).
[5] B. W. Zhang, D. Jasnow, and D. M. Zuckerman, J. Chem. Phys. 132, 054107 (2010).