AN EFFICIENT HYBRID STOCHASTIC-DETERMINISTIC SIMULATION TECHNIQUE FOR LIVING CELLS

Allocation: Illinois/169 Knh
PI: Zaida Luthey-Schulten[1]
Collaborators: David Bianchi[1], Tyler Earnest[1], Joseph Peterson[1], Michael Hallock[1]

[1]University of Illinois at Urbana-Champaign

MP

BI

EXECUTIVE SUMMARY

Stochasticity in gene expression is an important source of noise that can have profound effects on the fate of a living cell. The reactions for gene expression, feedback loops, and transport occurring within cells are typically described by Chemical Master Equations (CME). Sampling the CME using the Stochastic Simulation Algorithm (SSA) results in large computational costs as each reaction event is evaluated explicitly. To improve the computational efficiency of cell simulations involving high-particle-number systems, the authors have implemented a hybrid stochastic–deterministic (CME/ODE) method into the publicly available, GPU-based lattice microbes (LM) software suite, providing a convenient way to simulate complex cellular systems and interface with high-performance CME/RDME/ODE solvers. As a test of the implementation, the authors apply the hybrid CME–ODE method to the galactose switch in *Saccharomyces cerevisiae*, gaining a 10–50× speedup.

RESEARCH CHALLENGE

Many processes within living cells, especially gene expression, are characterized by low particle numbers and a high degree of randomness. The Chemical Master Equation (CME) and its spatially resolved analog, the Reaction–Diffusion Master Equation (RDME), are descriptions of cellular processes where the system is considered to follow a Markov jump process on the state space of particle numbers in time, capturing the discreteness of the particles and the random nature of individual chemical reactions. Gillespie's widely used Stochastic Simulation Algorithm (SSA) [1] provides an effective method for obtaining unbiased realizations of these Markov processes. This algorithm is limited by the fact that reaction events are accounted for explicitly by the SSA however, making simulations of highly reactive systems, where the time between reactions is small, computationally expensive. Highly reactive systems are characterized by large reaction propensities that can arise in the case of high copy numbers, such as metabolites in millimolar concentrations, and/or large rate constants (fast reactions). A challenging and typical scenario is when species participating in slow reactions interact with species involved in fast reactions, making the dynamics of the slow reactions dependent on the fast reactions. To alleviate the issues faced by the SSA for high particle number systems, many researchers have developed hybrid multiscale stochastic approaches [2–4] in which the highly reactive parts of the system are described by ordinary differential equations (ODE), and the slow reactive parts are described stochastically. Our hybrid method, along with an easy-to-use interface through LM [5] and pyLM [6], provides an effective way to study stochastic behavior in highly reactive systems.

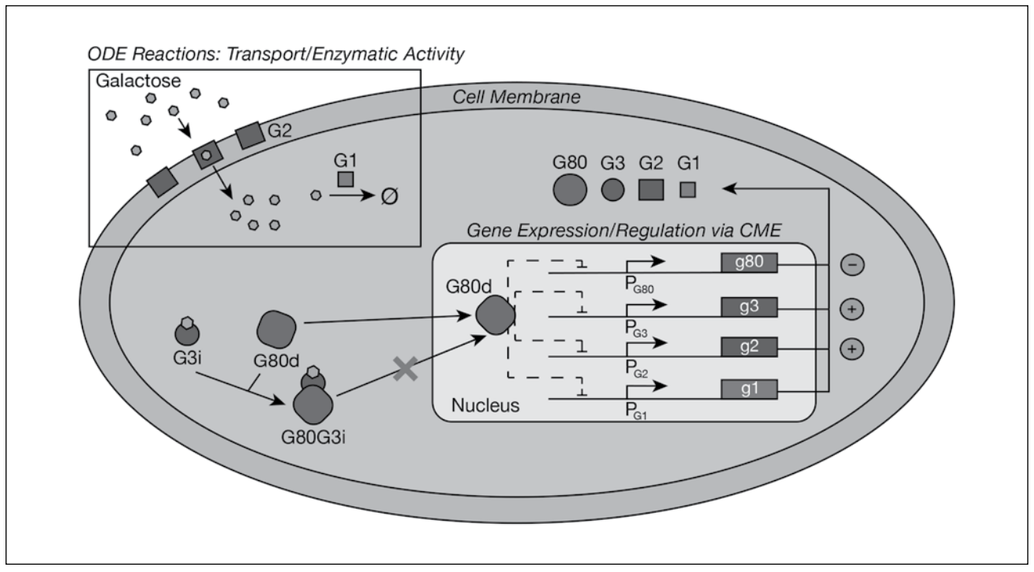

Figure 1: Schematic model of the galactose switch in yeast. The reactions depicted in the boxed area are simulated deterministically via an ODE solver, while those outside this region are simulated stochastically using the SSA.

METHODS & CODES

The galactose switch system, with its four feedback loops and millimolar galactose concentration, is separated into a regime of species whose reactions will be simulated stochastically and another whose reactions will be simulated deterministically (see Fig. 1). At the beginning of each timestep, the LSODA differential equation solver is updated with the species counts obtained from the stochastic regime (transcription, translation) simulated via the SSA, and then takes adaptive timesteps to evolve the high particle number species through time in the deterministic regime. At the conclusion of a timestep, the stochastic rates of reactions involving low particle number species interacting with high particle number species are updated with the species counts found by the ODE solver. The hybrid algorithm also communicates updated species counts generated from reactions in the CME regime to the ODE regime at this time. The optimal communication times between the stochastic and deterministic descriptions, as well as the timesteps for each method, need to be assessed to verify that the hybrid description accurately describes the stochastic dynamics, which often have great impact on the cell's behavior.

RESULTS & IMPACT

Such a CME–ODE partitioning works well for both bacterial and eukaryotic systems where stochastic effects are important. The partitioning typically improves the speed of the numerical simulations by a factor of 25–50, making it an indispensable tool for complex cell simulations with a large number of species types, cellular components, and high concentrations of metabolites (sugars, etc.) inside and outside the cell. Simulations enabled by this type of hybrid algorithm will allow researchers to study larger and more detailed systems, capturing the effects of reactions involving high particle count species such as metabolites, which have a crucial role in systems such as the genetic switch studied in this work. We have already used this hybrid approach to perform a spatially resolved RDME–ODE study (geometry presented in Fig. 2) of the galactose switch system, experiencing similar speedup to that seen in the CME implementation. The geometry of the yeast cell was derived from experimental cryo-electron tomography data through a fitting process and can be easily input using LM [7].

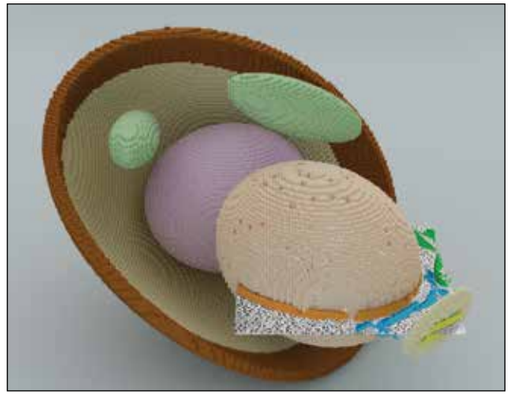

Figure 2: Experiment to LM—A spatially resolved yeast cell to be used in RDME simulation. The plane extending from the cell is the cryo-electron tomogram slice and segmentation data from which the simulation geometry was constructed [7].

WHY BLUE WATERS

Blue Waters was essential to generate thousands of replicate hybrid simulations over the simulation time of 750 minutes and a range of concentrations. Only then did we have sufficient data to make the results statistically reliable and to determine the optimal communication time. In the worst case scenario, the full CME simulations take nearly two days of wall-clock time, while the hybrid CME–ODE implementation often requires approximately 40 minutes. The response of the switch guided the setup for much more computationally costly RDME–ODE simulations on Blue Waters, which account for the spatial heterogeneous environment (nucleus, cytoplasm, membrane, etc.) of a cell.

PUBLICATIONS & DATA SETS

Bianchi, D., et al., Hybrid CME-ODE Method for efficient simulation of the galactose switch in yeast. *IET Systems Biology*, in press (2018), DOI:10.1049/iet-syb.2017.0070.

Earnest, T., et al., Challenges of Integrating Stochastic Dynamics and Cryo-electron Tomograms in Whole-Cell Simulations. *J Phys. Chem. B*, 121:15 (2017), pp. 3871–3881.