A HYBRID STOCHASTIC-DETERMINISTIC SIMULATION METHOD ENABLES FAST SIMULATION OF CELLULAR PROCESSES IN EUKARYOTES

Allocation: Illinois/169 Knh
PI: Zaida Luthey-Schulten[1]
Collaborators: David Bianchi[1], Joseph Peterson[1], Tyler Earnest[2], Michael Hallock[1]

[1]University of Illinois Urbana-Champaign
[2]National Center for Supercomputing Applications

MP

EXECUTIVE SUMMARY

Stochasticity in transcription is an important source of noise that can have a profound effect upon the fate of a living cell. In recent years, we have seen the advent of a community of researchers interested in performing stochastic simulations of large biological systems (e.g., millions of particles). The Lattice Microbes (LM) software suite and its pyLM problem-solving environment provide a convenient way to set up simulations of complex biological systems. However, simulations of large systems performed using the exact Stochastic Simulation Algorithm (SSA) to solve the Chemical Master Equation (CME) are computationally expensive. To alleviate this issue, we have implemented a hybrid CME-ODE (ordinary differential equations) method for LM similar to previous mixed methodologies. Our robust hybrid implementation gives protein distributions and species dynamics similar to the pure SSA and can run approximately 50–100X faster in a simulation of the well-studied galactose switch in S. cerevisiae (yeast) containing millimolar particle concentrations.

RESEARCH CHALLENGE

Many processes within living cells, especially gene expression, are characterized by low particle numbers and a high degree of randomness. 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 SSA method [1] provides an effective technique for obtaining unbiased realizations of these Markov processes.

However, this algorithm is limited because reaction events are accounted for explicitly by the SSA, 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 [2,3,4] in which the highly reactive parts of the system are described by 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.
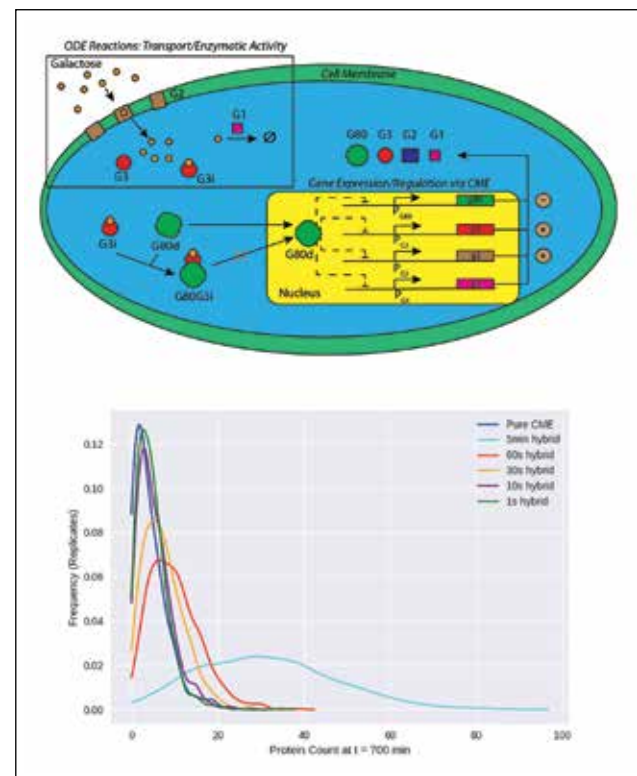

Figure 1: (top) The S. cerevisiae galactose switch system with ODE/CME separation. The reactions depicted (boxed area) are simulated continuously; those outside this region are simulated stochastically. (bottom) Distribution at steady state of the G3i species. As the communication timestep is decreased from 5 minutes to 1 second the distributions given by the hybrid algorithm more closely match pure SSA results.

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 (Fig. 1). At the beginning of each timestep, the differential equation solver (DES) is updated with the species counts obtained from the stochastic regime (transcription, translation) simulated via the SSA. The DES 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. At this time, the hybrid algorithm also communicates updated species counts generated from reactions in the CME regime to the ODE regime. The optimal communication times between the stochastic and deterministic descriptions, as well as the timesteps for each method, need to be assessed (Fig. 1) 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. Partitioning typically improves the speed of the numerical simulations by a factor of 50–100, 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 here. We have already used this hybrid approach to perform a spatially resolved RDME-ODE study (Fig. 2) of the galactose switch system, experiencing similar speedup to what is seen in the CME implementation.

WHY BLUE WATERS

Blue Waters was essential to generate over 1,000 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 with a communication time of one second requries less than 30 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 the yeast cell.
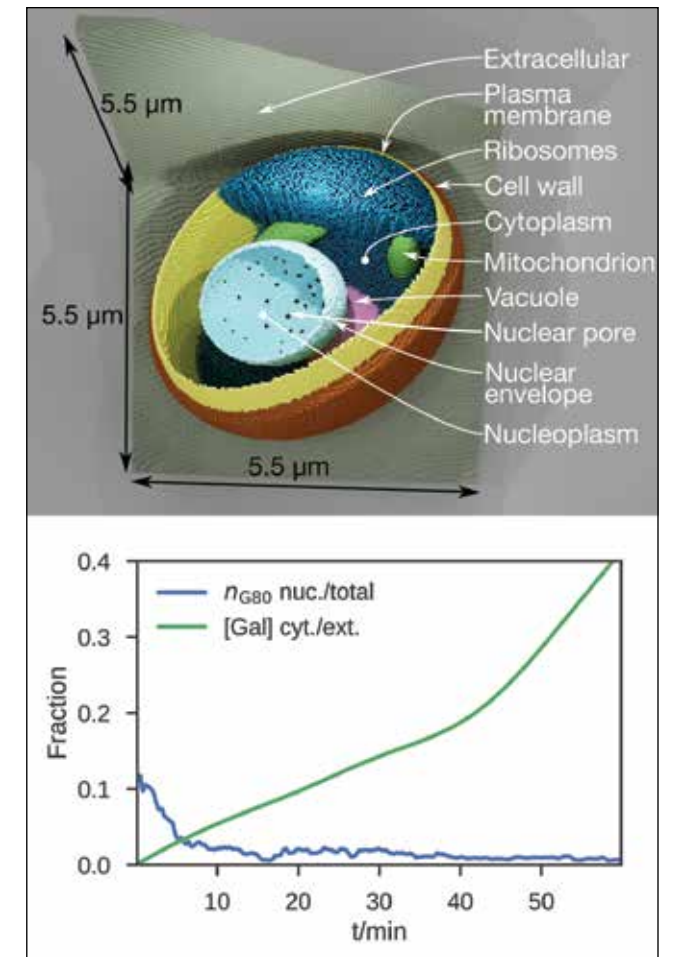

Figure 2: (top) A visualization of the cell compartmentalization that is imposed in a hybrid RDME-ODE simulation, with structural insight taken from tomograms. (bottom) The decrease in the amount of transcription factor present in the nucleus as the galactose concentration increases, matching expectations from experimental results.