Flexible state-process framework for simulating molecular systems

The goal of this project is to explore cellular processes that span a range of timescales using simulations in state-process models.

Anyone thinking about a natural phenomenon has a model in mind. It can be explicitly expressed in mathematical notation, or remain implicit during reasoning, testing, and communicating. Most often, a biological model is represented by scribblings on the whiteboard, with nodes corresponding to molecules, complexes, and locations, connected by arrows representing activation, repression, relocation, degradation, etc. There is no agreed-upon graphical language for these pictures, which leaves the interpretation dependent on many implicit qualifications. As language shapes the thoughts and ideas we express, the types of models we use frame our thinking and experimental design.

The goal of this project is to explore cellular processes that span a range of timescales using simulations in state-process models. The state X_t of the cellular system at time t is described by N state variables x_{t,n}, so X_t = (x_{t,1}, ..., x_{t,N}). For example, a state could be the number of certain mRNA molecules, number of ATP molecules, number of PolII-bound promoters of a gene, or cell cycle stage. The state can have associated uncertainty, and be represented by a distribution that can be used to calculate average values and higher moments.

The state changes due to action of M cellular processes F = (f_1, ... , f_M). Every process f_m takes in the current cellular state X_t, and calculates the change \Delta_{t,m} to all the state variables impacted by it over time dt. The processes capture either a single biochemical reaction (a low-level process with short timescale such as transcription factor binding), or more complicated ongoings, such as TCA cycle, or replication of all DNA. The definition reduces to a standard time-difference model using a graph Laplacian (e.g. Gunawardena 2013), \frac{dX_t}{dt} = L X_t, in a special case, but does not require linearity in the state, and can make use of uncertainty in the variables. As a downside, analytic solutions are likely not obtainable in general. Thus, the main goal of this model is accurate and intuitive simulation.
This model formalism enables intuitive and efficient calculation. First, evolution of the model over time can be captured with a message passing algorithm, where processes and states interact via sending information about their statistics. The messages can be passed at timescales that depend on local conditions - if states fluctuate a lot, and the fluctuation has large downstream consequences, the timescale can be tuned to be shorter. Finally, several processes operating at similar timescales could be combined into a single one with matching input/output characteristics, allowing simpler descriptions, and natural compositions.