

# Stochasticity in Biochemical Reaction Networks

## September 25-27, 2009

### MEALS

*All meals are included with the workshop (starting with dinner on Friday, 25 September to lunch on Sunday, 27 September).*

Breakfast (Buffet): 7:00–9:30 am, Sally Borden Building

Lunch (Buffet): 11:30 am–1:30 pm, Sally Borden Building

Dinner (Buffet): 5:30–7:30 pm, Sally Borden Building

Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall (*included in workshop*)

Please remember to scan your meal card at the host/hostess station in the dining room for each meal

### MEETING ROOMS

**All lectures are held in Max Bell 159. LCD projector, overhead projectors and blackboards are available for presentations** *Please note that the meeting space designated for BIRS is the lower level of Max Bell, Rooms 155-159. Please respect that all other space has been contracted to other Banff Centre guests, including any Food and Beverage in those areas.*

## SCHEDULE

### Friday

- 16:00** Check-in begins (Front Desk - Professional Development Centre - open 24 hours)  
Lecture rooms available after 16:00
- 17:30** Dinner
- 19:30** Informal gathering in 2nd floor lounge, Corbett Hall (if desired)  
Beverages and small assortment of snacks available in the lounge on a cash honour-system.

### Saturday

- 7:00-9:00** Breakfast
- 8:45** Welcome
- 9:00** Kyung-Hyuk Kim: *Sensitivity Analysis and Module Interface Conditions on Stochastic Gene Circuits*  
Mary Dunlop: *Regulation Revealed by Correlations in Gene Expression Noise*  
Aleksandra Walczak: *Information processing in small gene regulatory networks and cascades*
- 10:30** Coffee Break, 2nd floor lounge, Corbett Hall
- 11:00** Ilya Nemenman: *Mesoscopic statistical properties of complex, multi-step biochemical reactions*  
Jim Faeder: *Network-free simulation of rule-based biochemical models*
- 12:00** Lunch
- 13:30** Arjun Raj: *Nature, nurture or just dumb luck: from single molecules to cell fate*  
Narendra Maheshri: *Noise induces bimodal gene expression patterns in positive feedback loops without cooperativity*
- 14:30** Coffee Break, 2nd floor lounge, Corbett Hall
- 15:00** David Soloveichik: *Computability Properties of Chemical Reaction Networks*  
Marc Riedel: *The Computer-Aided Synthesis of Modular Stochastic Biochemistry*
- 16:00** Discussion Break
- 16:30** William Shih: *Self-assembly of DNA into nanoscale three-dimensional shapes*  
Henry Hess: *Herding Nanotransporters: Localized Activation via Release and Sequestration of Control Molecules*
- 17:30** Reception with BIRS Board of Directors (Max Bell Lounge)
- 18:30** Dinner

### Sunday

- 7:00-9:00** Breakfast
- 9:00** Gregor Neuert: *Information Processing in Individual Cells: A Single Molecule Approach to Systems Biology*  
Brian Munsky: *Listening to the Noise: Random Fluctuations Reveal Gene Network Parameters*  
Michael Samoilov: *Existence of Qualitative Stochastic Effect in Large Molecular Systems*
- 10:30** Coffee break, 2nd floor lounge, Corbett Hall (Checkout by 12 noon)
- 11:15** David Thorsley: *Observers for Stochastic Chemical Kinetics*  
Sotiria Lampoudi: *Space in Stochastic Simulation of Intracellular Kinetics*  
Michael Chevalier: *An exact CME/SSA formulation for separating timescales in stochastic biological networks*
- 12:45** Lunch
- 1:30** Discussion and Farewell

\*\* 2-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 15:00 on Sunday, although participants are still required to checkout of the guest rooms by 12 noon. There is no coffee break on Sunday afternoon, but self-serve coffee and tea are always available in the 2nd floor lounge, Corbett Hall. \*\*

# Stochasticity in Biochemical Reaction Networks

## September 25-27, 2009

### ABSTRACTS (in alphabetic order by speaker surname)

Speaker: **Michael Chevalier** (University of California, San Francisco)

Title: *An exact CME/SSA formulation for separating timescales in stochastic biological networks*

Abstract: Noise and stochasticity are fundamental to biology and derive from the very nature of biochemical reactions where thermal motion of molecules translates into randomness in the sequence and timing of reactions. This randomness leads to cell-cell variability even in clonal populations. Stochastic biochemical networks are typically modeled as continuous-time, discrete-state Markov processes whose probability density functions evolve according to a chemical master equation (CME). The CME is not solvable but for the simplest cases, and one has to resort to kinetic Monte Carlo techniques to simulate the stochastic trajectories of the biochemical network under study. A commonly used such algorithm is the stochastic simulation algorithm (SSA). Because it tracks every biochemical reaction that occurs in a given system, the SSA presents computational difficulties especially when there is a vast disparity in the timescales of the reactions or in the number of molecules involved in these reactions. This is common in cellular networks, and many approximation algorithms have evolved to alleviate the computational burdens of the SSA. Here, we present a rigorously derived modified CME framework based on the partition of a biochemically reacting system into restricted and unrestricted reactions. Although this modified CME decomposition is as analytically difficult as the original CME, it can be naturally used to generate a hierarchy of approximations at different levels of accuracy. Most importantly, some previously derived algorithms are demonstrated to be limiting cases of our formulation. Time permitting, we will also discuss some spatial effects in stochastic biological networks.

Speaker: **Mary Dunlop** (Lawrence Berkeley National Laboratory)

Title: *Regulation Revealed by Correlations in Gene Expression Noise*

Abstract: Gene regulatory interactions are context-dependent, active in some cellular conditions but not in others. I will discuss a method for determining when a regulatory link is active given temporal measurements of gene expression. Correlations in time series data are used to determine how genes influence each other and their causal relationships. Natural stochastic noise is shown to aid in the process of network identification by perturbing the expression of genes; the speed and direction at which the noisy signal propagates shows how the network is connected. I will present results from a three-color synthetic gene circuit imaged with time-lapse single-cell microscopy. By using a synthetic gene circuit with well-characterized connections we were able to validate our approach by reproducing known system characteristics. In addition, I will discuss results where a synthetic reporter construct was used to measure gene-expression in a natural regulatory network responsible for galactose metabolism.

Speaker: **Jim Faeder** (University of Pittsburgh)

Title: *Network-free simulation of rule-based biochemical models*

Abstract: Signalling in cells generally involves protein-protein interactions, which can produce myriad protein complexes. Such protein-protein interactions can be represented compactly and precisely using graphical reaction rules, which can be processed automatically to obtain a chemical reaction network. However, reaction networks implied by typical sets of rules are often too large for conventional simulation procedures to handle. To address this challenge, we have developed a kinetic Monte Carlo method that can take advantage of a rule-based model specification. Rules are used directly to advance a simulation, thus avoiding the computationally expensive step of generating the underlying chemical reaction network implied by the rules. Unlike previously proposed methods that adaptively generate species and reactions in response to network activity, the method is not overwhelmed when the likelihood of encountering new species each time a reaction fires becomes high. The method is illustrated by using it to characterize the interaction of a trivalent ligand with a bivalent cell-surface receptor. The results of the simulations suggest formation of extremely large receptor aggregates under typical experimental conditions.

Speaker: **Henry Hess** (University of Florida / Columbia University)

Title: *Herding Nanotransporters: Localized Activation via Release and Sequestration of Control Molecules*

Abstract: A challenge for nanotechnology is the dynamic and specific control of nanomachines by the user. Molecular shuttles, consisting of cargo-binding microtubules propelled by surface-immobilized kinesin motor proteins, are an example of a nanoscale system that ideally can be selectively activated at programmable locations and times. Here we discuss a biomimetic solution where activating molecules are delivered locally via photolysis of a caged compound and subsequently sequestered in an enzymatic network. The controlled sequestration of the activator not only creates a rapid deactivation when the stimulus is removed but also sharpens the concentration profile of the rapidly diffusing activator. This improvement comes at the expense of a reduced efficiency in the utilization of the activator molecules, suggesting that these nanosystems are most efficiently addressed as a swarm rather than as individuals. Our work represents a step toward transferring the cellular control strategies of molecular activation to bionanotechnology.

Speaker: **Kyung-Hyuk Kim** (University of Washington)

Title: *Sensitivity Analysis and Module Interface Conditions on Stochastic Gene Circuits*

Abstract: In this research we describe an extension of the deterministic metabolic control analysis (MCA) to the stochastic regime. We introduce stochastic sensitivities for mean and covariance values of reactant concentrations and reaction fluxes and show that there exist MCA-like summation theorems among these sensitivities. The summation theorems for flux variances and the control distribution of the flux variances is shown to depend on the size of the measurement time window ( $\epsilon$ ) within which reaction events are counted for measuring a single flux. It is found that the degree of the  $\epsilon$ -dependency can become significant for processes involving multi-time-scale dynamics. We also propose a systematic way to control fluctuations of reactant concentrations while minimizing changes in mean concentration levels by applying the stochastic sensitivities.

Gene regulatory circuits show significant stochastic fluctuations in their circuit signals due to the low copy numbers of transcription factors. When a gene circuit component is connected to an existing circuit, the dynamic properties of the existing circuit can be affected by the connected component. We show that the output signal of the existing circuit can change significantly in its noise component due to the interaction with the downstream circuit. More specifically, the output signal noise shows longer correlations compared to the case when disconnected, or, equivalently, the noise spectrum has a shorter range of frequency response. We define the relative change in the correlation time by stochastic retroactivity, which will be shown to be directly related to the retroactivity defined in the deterministic framework by del Vecchio et al. This provides an insight on how to observe the retroactivity, by investigating stochastic fluctuations in gene expression levels, more specifically, an autocorrelation function of the transcription factor expression levels.

Speaker: **Sotiria Lampoudi** (University of California, San Diego)

Title: *Space in Stochastic Simulation of Intracellular Kinetics*

Abstract: The Stochastic Simulation Algorithm (SSA) of Gillespie (1976) has become the golden standard for the stochastic simulation of intracellular kinetics, and the literature is brimming with accelerated, approximate and derivative algorithms. In this talk I will present recently published theoretical and numerical work which addresses two issues that arise from the application of the SSA to biological systems in particular: molecular crowding, and simulating not well-mixed systems. I will tackle the following questions: 1) how does the volume excluded by the reactants in a crowded setting impact the propensities of the SSA (J Comput Phys 2009)? and 2) how can the Inhomogeneous SSA (i.e. the extension of the SSA to a not well-mixed setting), which is a prohibitively expensive algorithm to apply to most systems, be made computationally tractable (J Chem Phys 2009)?

Speaker: **Narendra Maheshri** (Massachusetts Institute of Technology)

Title: *Noise induces bimodal gene expression patterns in positive feedback loops without cooperativity*

Abstract: Positive feedback is a common network motif in gene regulatory networks that is widely recognized to lead to bistability and, as a consequence, to hysteresis and switch-like responses. A non-linear, cooperative promoter response provides the necessary ingredient to generate bistability in deterministic descriptions of positive feedback. Using a synthetic system, we show experimentally that positive feedback is capable of inducing a bimodal, switch-like response with non-cooperative feedback, even when the underlying deterministic dynamics do not admit bistability. In accordance with theoretical models, the bimodal response requires the promoter within the feedback loop to be noisy, with infrequent, large bursts of expression. In addition, the transcription factor (TF) involved in the feedback loop has to be short-lived. Using a stochastic model and experimentally measured /in vivo/ parameters of the promoter response in the absence of feedback, we can quantitatively describe the feedback response. We also find that multiple TF binding sites in a promoter can be important for the bimodal response not because of molecular cooperativity in TF binding, but because of increased noise in the promoter. Because many promoters possess multiple binding sites and many TFs are unstable, positive feedback loops in many gene regulatory networks may exhibit bimodal responses, but not necessarily because of deterministic bistability as is commonly thought.

Speaker: **Brian Munsky** (Los Alamos National Laboratory)

Title: *Listening to the Noise: Random Fluctuations Reveal Gene Network Parameters*

Abstract: The cellular environment is abuzz with noise originating from the inherent random motion of reacting molecules in the living cell. In this noisy environment, clonal cell populations exhibit cell-to-cell variability that can manifest significant phenotypic differences. Noise induced stochastic fluctuations in cellular constituents can be measured and their statistics quantified using flow cytometry and other single cell measurement techniques. We show that these random fluctuations carry within them valuable information about the underlying genetic network. Far from being a nuisance, the ever-present cellular noise acts as a rich source of excitation that, when processed through a gene network, carries its distinctive fingerprint that encodes a wealth of information about that network. We demonstrate that in some cases the analysis of these random fluctuations enables the full identification of network parameters, including those that may otherwise be difficult to measure. We use theoretical investigations to establish experimental guidelines for the identification of gene regulatory networks, and we apply these in the experimental identification of different regulatory mechanisms in E. coli and yeast.

Speaker: **Ilya Nemenman** (Emory University)

Title: *Mesoscopic statistical properties of complex, multi-step biochemical reactions*

Speaker: **Gregor Neuert** (Massachusetts Institute of Technology)

Title: *Information Processing in Individual Cells: A Single Molecule Approach to Systems Biology*

Abstract: How cells sense their environment using signal transduction pathways and respond to environmental changes by regulating gene expression is a key problem in systems biology. Our research focuses on the high-osmolarity glycerol (HOG) pathway, which is one of the mitogen-activated protein kinase (MAPK) pathways in *Saccharomyces cerevisiae* yeast cells. During the last few decades, the components and regulatory network of this pathway have been elucidated via genetic and biochemical assays performed on large populations of yeast cells. However, surprisingly little is known about the dynamics of signal transduction and its regulation of gene expression in individual cells. We have found that signal transduction dynamics and intensity in the HOG - pathway in single cell is homogeneous but the subsequent gene expression of STL1, a gene that encodes for a glycerol proton symporter of the plasma membrane, is heterogeneous. In addition, we found that gene expression pattern is bi-modal leaving a constant fraction of cells non-responsive independent of the strength of the imposed signal. Furthermore, we found through single cell time-lapse microscopy experiments that gene expression activation after repetitive signal transduction is random and memory less. To identify the mechanism that can explain these findings, we are using a combination of quantitative single molecule experiments, genetics and mathematical modeling. Using genetics, we identified a transcription factor, which modulate the fraction of ON and OFF cells but not their gene expression levels. In addition to the genetic approach we used the quantitative single molecule RNA FISH technique to determine the absolute number of endogenous STL1 mRNA transcripts during transcription. These Results from these experiments have been compared directly to single molecule based model approaches which has helped us to formulate an intuitive model that captures all the above mentioned findings.

Speaker: **Arjun Raj** (Massachusetts Institute of Technology)

Title: *Nature, nurture or just dumb luck: from single molecules to cell fate*

Abstract: Gene expression, the biochemical process by which a cell's genetic code is read out to produce proteins, is a fundamentally stochastic process. One consequence is that even genetically identical populations of cells in homogenous environments can often display significant cell-to-cell variability in the numbers of mRNAs and proteins. This raises a couple of questions. Can cells exploit this randomness in the execution of the genetic code for their own benefit? Conversely, do cells reduce the impact of variability in order to produce reliable outcomes in other contexts? We have developed an method that allows for the detection of individual mRNA molecules in a host of organisms to help us answer these questions by providing us with very sensitive measures of gene expression in individual cells. In particular, we have explored the impact of variability in gene expression on the process of development by studying the gene regulatory network responsible for gut formation during early embryonic development in *C. elegans*. We found that the normal gut development pathway is remarkably robust, but this robustness can be destroyed by mutations to a single gene that result in wildly varying embryonic fates. We have shown that these different fates result from the variable expression of a key upstream regulator in the rewired mutant gut pathway. These results suggest that redundancy in developmental gene networks can serve to mask and buffer otherwise hidden sources of gene expression variability.

Speaker: **Marc Riedel** (University of Minnesota)

Title: *The Computer-Aided Synthesis of Modular Stochastic Biochemistry*

Abstract: Our research encompasses projects in the design and verification of digital circuits as well as in synthetic and computational biology. A broad theme is the application of computational expertise from the former (circuit design) to analysis and design problems in the latter (biology). A specific theme that cuts across both domains is constructing and deconstructing probabilistic behavior. In the biological realm, we are designing biochemical pathways that produce different combinations of molecular types according to programmable probability distributions. This gives us the ability to fine-tune a probabilistic response – akin to hedging with a portfolio of investments. In the engineering realm, we are designing digital circuits that process zeros and ones probabilistically. This is a promising strategy for coping with the randomness that occurs due to noise and glitches as circuit components are scaled down to nanometers in size.

Randomness is inherent to biochemistry: at each instant, the sequence of reactions that fires is a matter of chance. Some biological systems exploit such randomness, choosing between different outcomes stochastically. We have developed a computer-aided design tool called BAMBI (Brian’s Automated Modular Biochemical Instantiator) that synthesizes biochemical reactions with stochastic behavior. Given a library of parts it selects biochemical reactions that produce different combinations of molecular types according to a specified probability distribution. The response is precise and robust to perturbations. Furthermore, it is programmable: the probability distribution is a function of the quantities of input types. The method is modular and extensible. We discuss strategies for implementing various functional dependencies: linear, logarithmic, exponential, etc.

Speaker: **Michael Samoilov** (QB3/University of California, Berkeley)

Title: *Existence of Qualitative Stochastic Effect in Large Molecular Systems*

Abstract: The time evolution of biochemical molecular systems at the macroscopic level is most commonly modeled by using reaction rate continuous-deterministic classical chemical kinetic equations (CCK). These can be derived via a series of approximations from the more accurate microscopic/mesoscopic descriptions of discrete-stochastic molecular collisions and reaction probabilities, such as offered by the chemical master equation (CME). It is commonly assumed that significant qualitative deviations in CCK predictions from the behaviors otherwise described by the CME arise either at low molecular counts or transiently before the stationary limit is reached. This talk will discuss potential mechanisms for violating these assumptions and trace the kinetic origins of such deviant effects.

Speaker: **William Shih** (Harvard Medical School)

Title: *Self-assembly of DNA into nanoscale three-dimensional shapes*

Abstract: I will present a general method for solving a key challenge for nanotechnology: programmable self-assembly of complex, three-dimensional nanostructures. Previously, scaffolded DNA origami had been used to build arbitrary flat shapes 100 nm in diameter and almost twice the mass of a ribosome. We have succeeded in building custom three-dimensional structures that can be conceived as stacks of nearly flat layers of DNA. Successful extension from two-dimensions to three-dimensions in this way depended critically on calibration of folding conditions. We also have explored how targeted insertions and deletions of base pairs can cause our DNA bundles to develop twist of either handedness or to curve. The degree of curvature could be quantitatively controlled, and a radius of curvature as tight as 6 nanometers was achieved. This general capability for building complex, three-dimensional nanostructures will pave the way for the manufacture of sophisticated devices bearing features on the nanometer scale.

Speaker: **David Soloveichik** (California Institute of Technology)

Title: *Computability Properties of Chemical Reaction Networks*

Abstract: We consider the behavior of Stochastic Chemical Reaction networks as a form of computation, and survey connections between SCRNs and computational models such as Boolean Logic Circuits, Vector Addition Systems, Petri nets, Primitive Recursive Functions, Register Machines, and Turing Machines. A theme to these investigations is the thin line between decidable and undecidable questions about SCRN behavior. That SCRNs are naturally capable of complex computational behavior helps explain why simulating them is computationally difficult.

Speaker: **David Thorsley** (University of Washington)

Title: *Observers for Stochastic Chemical Kinetics*

Abstract: The state of a stochastic chemical kinetic model is a vector where each element is the molecular population of one of the species in the biochemical process. Using such models, we can make predictions regarding the distributions of dynamic behaviors of cellular processes that can be tested using single-cell experiments. However, it is not experimentally possible to completely observe the evolution of the state vector as it varies with time. Time-lapse movies collected using single-cell fluorescence microscopy allow the experimenter to make estimates of the dynamic populations of a few fluorescent species. Observations of dynamic behavior are not made continuously but instead at intermittent time points. A well-designed experiment ensures that these limited observations provide enough useful information to draw conclusions about the behavior of unobservable species. In control theory, the standard approach to dealing with the problem of estimating a dynamically changing state from limited observations is to construct an observer. An observer is a system that receives measurements from a system being monitored and computes an estimate of the monitored system's state. In this work, we develop an observer structure specialized for systems described using the formalism of stochastic chemical kinetics and specialized for quantitative analysis of time-lapse movies made using fluorescence microscopy. We show how observers can be applied to the following problems of state estimation, diagnosis of events of interest, and hypothesis testing and parameter estimation.

Speaker: **Aleksandra Walczak** (Princeton University)

Title: *Information processing in small gene regulatory networks and cascades*

Abstract: Many of the biological networks inside cells can be thought of as transmitting information from the inputs (e.g., the concentrations of transcription factors or other signaling molecules) to their outputs (e.g., the expression levels of various genes). On the molecular level, the relatively small concentrations of the relevant molecules and the intrinsic randomness of chemical reactions provide sources of noise that set physical limits on this information transmission. Given these limits, not all networks perform equally well, and maximizing information transmission provides a candidate design principle from which we might hope to derive the properties of real regulatory networks. As a starting point, I will consider the simple case of one input transcription factor that controls many genes. I will discuss the properties of these specific small networks that can transmit the maximum information. Concretely, I will show how the form of molecular noise drives predictions not just of the qualitative network topology but also the quantitative parameters for the input/output relations at the nodes of the network. Lastly, I will generalize the considerations of information flow to gene regulatory cascades by introducing the spectral method for computation of the joint probability distribution over all species in a biological network. The spectral method finds the distribution using a fast and accurate method that exploits the natural basis of the uncoupled problem from the same class. I will show, that for threshold regulation, a cascade of strong regulations converts a unimodal input to a bimodal output, that multimodal inputs are no more informative than bimodal inputs, and that a chain of up-regulations outperforms a chain of down-regulations.