Stony Brook University

Stony Brook University

Non-adiabatic Dynamics of Gene Regulatory Network

A Dissertation presented

by

Cong Chen

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

December 2015

Stony Brook University

The Graduate School

Cong Chen

We, the dissertation committe for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Jin Wang - Dissertation Advisor
Professor, Departments of Chemistry and Physics

Thomas Weinacht - Chairperson of Defense
Professor, Department of Physics and Astronomy

Dmitri Averin
Professor, Department of Physics and Astronomy

Xiaolin Li
Professor, Department of Applied Math and Statistics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Non-adiabatic Dynamics of Gene Regulatory Network

by

Cong Chen

Doctor of Philosophy

in

Physics

Stony Brook University

2015

Gene regulatory network is a mathematical model of gene expression and regulation in cell environment. The concept of network comes from network in math and computer science. In gene network each node is a gene that has different expression level and links between nodes represent direct regulation (activation or repression). Gene network is stochastic with intrinsic noise from biochemical reactions involved and extrinsic noise from environment. It is non-equilibrium with frequent matter/energy exchange and active entropy production. One crucial character of gene network is the involvement of multiple timescales: timescale of protein synthesis/degradation and timescale of regulation processes. Conventional studies are concentrated at adiabatic limit where regulation processes are much more frequent than protein synthesis/degradation and adiabatic approximation is valid. We explore non-adiabatic dynamics of gene network by develop a mapping from N-dimensional protein concentration space to 2N extended space using similarity to quantum mechanics and path integral. We applied our theory to self activator which is the simplest network motif. Our theory is able to explain steady states at different adiabaticity and demonstrates non-equilibrium properties like eddy current. We also studied relationship between cancer heterogeneity and

non-adiabatic dynamics of core cancer network. In non-adiabatic regime, the cancer network shows alternation of phenotypic states, weaker stability, optimal transition rate and diversity of transition paths. This suggests possible source of cancer heterogeneity from non-adiabatic dynamics of core cancer network.

To my parents.

Contents

List of Figures

Acknowledgements

It is a good place to show my gratitude to Professor Jin Wang, my thesis advisor. He introduced me into the active research field of gene network and always inspired me with creative ideas and passionate working attitude. This work would not have been possible withou his support, assistance, frequent discussion and encouragement.

I would like to thank my research group members. Especially Dr. Chunhe Li who helped me with numerical calculation and programing, Dr. Haidong Feng and Dr. Wei Wu for helpful discussion, Dr. Yuan Yao who constantly encouraged me. Your friendship always warmed my heart.

I would also like to thank Jie Ma, Chengjian Wu, Xugang He, Adrian Soto, Saad Nadeem, Si Wen and many other friends in physics department and in stony brook university. It is a special journey and I'm lucky to have met you.

I'm grateful to the Brownworths for many years' friendship and support.

Thank you goes to my committee members for taking time out of their busy schedule, read my thesis and come to my defense.

Last but not least, I'm gratefull to my parents Jinying Zhang and Sheng Chen for their endlee support and love all these years.

Chapter 1

Introduction

1.1 Gene Regulatory Network

[1] [2]

A gene is a piece of DNA sequence that encodes a functional protein. Within a cell, proteins can be structural and accumulate, or they can be enzymes that catalyse specific reactions. There are thousands of genes in a single cell, the complex functions of the cell is largely determined by the expression of different genes. For example, yeast has over 5000 genes. For the fermentation process during which yeast processes sugar into ethanol, the expression level of over 2000 genes are changed. What is the mechanism for different genes to work coorperately during the process? Human has about 20000-25000 genes. For a single cell in human, only 20% of all human genes are expressed at any given time. What is the mechanism that determines certain genes are expressed while certain genes are repressed

[1] Chapter 3 is largely a reprint of a published paper coauthored with my advisor Professor Jin Wang and Kun Zhang, Haidong Feng, Masaki Sasai:
"Multiple coupled landscapes and non-adiabatic dynamics with applications to self-activating genes"
Physical Chemistry Chemical Physics, 2015, DOI: 10.1039/C5CP04780C

[2] Chapter 4 is largely a reprint of a submitted paper coauthored with Professor Jin Wang to Scientific Reports and is currently under review

in the same cell environment?

The answer to both questions is the gene regulation mechanism. It is now understood that the expression of a gene (whose transcription product is a protein or an RNA) is not determined solely by the gene itself. It can be activated or repressed by another gene. The regulation usually takes the form of binding/unbinding of a regulatory protein that is the translation product of the regulatory gene to a piece of DNA sequence on the target gene that controls the transcription process.

The fact that a gene can be activated or repressed by another gene makes the whole system a network. Within the network, each node represents a protein, mRNA or protein/protein complexes that is the transcription/translation product of a gene. The links or edges between nodes are activation or repression reactions. The evolution of the network controls the expression level of protein and RNA products.[1, 2, 3].

Gene network is essential to the biological behaviors of the cell. In a living organism, though all the cells share the same genome, they can show very different phenotypes. Different genes can be activated and as a result cells differentiate differently and have different fates.[4] This cell differentiation process is controlled by gene network.[5] During the differentiation process, the network has to be stable so that differentiated cells of the same type can have consistent gene expression levels and stable functionality.[6, 7] There's also uncertainty associated with the network so that mutations can happen and biological evolution is possible. Gene network responds to external environment. In example, when yeast cells are placed in sugar solution, the high sugar concentration will trigger the fermentation network to produce enzymes to process sugar into alcohol.[1] Similar behaviors are found in multicellular animals. This kind of behaviors is essential for cells to survive and evolve in different environments.[8, 9]

Gene network has vast implications in biology and clinical therapy. For example in one of the most fundamental problems in biology, the cell fate problem. Stem cells that share the same genome will differentiate into different cells with different functionality. The differentiation process is controlled by gene network. Recently there has been exciting development in this research field. Researchers are able to transform a differentiated cell from animal into a multi-potent cell that has the ability to re-differentiate.[10, 11]. This makes reprograming and control of differentiation network possible. It has the future potential to make regenerative medicine.[12] Another important implication is the cancer core network that controls the development of cancer. It is now understood cancer is caused by collective gene-gene interactions as well as single gene mutation.[13, 14, 15, 16, 17] From the gene network point of view, cancer can be seen as a biological state of gene regulatory network that is hidden deep under the complex network dynamics. Cancer state is not accessible in normal situations but can be reached in rare cases. The network point of view refreshes our knowledge about cancer and provides new ideas for cancer therapy.

To study gene network, on one hand we need to study the structure of the network through wiring diagram. The wiring diagram of a network provides a map of all regulatory interactions. For example we know a certain protein won't have a high concentration if its repressor is present, or will likely have a high concentration if its activator is present. For a toggle switch with two genes that repress each other, there will not likely be a stable state where both genes are activated at the same time.[18] A large network usually contains many small network motifs.[4] For example, one gene self-activator/repressor or two gene toggle switch.[19, 20] In other words, smaller network motifs are building blocks of larger network. It's important for us to understand behavior of small network motif

as well as structure of larger networks.

However, the wiring diagram of the network is not the whole story. To get a full understanding of gene network we need to study its dynamics as well. To study network dynamics, first of all we need to consider all bio-chemical reactions involved explicitly. Usually they include the protein synthesis/degradation and the regulatory reactions. One typical regulation reaction is through binding/unbinding of one regulatory protein or protein complexes to the promotor site of the target gene.[21] As we will see different reaction rates will result in different evolution results. Besides bio-chemical reaction rates, there are several factors about gene network we need to consider. Here we give a brief introduction.

1.2 Properties Of Dynamics

One of the major task in studying gene network dynamics is to identify all possible biological states that are stable steady states of gene network with different gene expression levels. Gene expression level is measured by corresponding protein copy number or concentration. Gene network dynamics, when considered in phase space composed of protein concentrations, is a multi-dimensional thermodynamic system. This thermodynamic system is different from conventional physical thermodynamic systems in several ways.

First of all. It is an open system that has frequent energy, matter and information exchange with outer cell environment. With frequent exchange of matter, energy and information, gene network is an open and non-equilibrium system.[22] As we shall see in the self activator and cancer core network example later, even after long evolution the system reaches steady state with stable probability distribution, there still can be a non-trivial flux that breaks detailed balance.[22, 23] As a result of system be-

ing non-equilibrium, the kinetic path from one stable biological state to another is irreversible. The forward and backward transition following the same path will not have equal probability. Furthermore, the optimal transition path doesn't follow the gradient of a potential function but drifted by the non-gradient steady state flux. This makes the forward and backward optimal transition paths different. The fact that forward and backward optimal paths go through different intermediate states can have implications in many biological systems and clinical therapy.[24] The non-equilibriumnes can be measured quantitatively by entropy production rate.

Second of all, gene network is a stochastic system with intrinsic and extrinsic fluctuations. In cell environment, the copy number of proteins is limited with typical number of a few hundreds.[25] The intrinsic fluctuation from bio-chemical reactions involved, though can be ignored in bulk chemical system when molecule numbers are large, needs to be considered explicitly in gene network and plays an important role in gene network evolution.[26] On the other hand, the extrinsic fluctuation from biological and thermal environment is important too. Intrinsic fluctuation and extrinsic fluctuation have different properties. Intrinsic fluctuation is 'local' meaning it depends on current protein copy numbers and gene network state. Extrinsic fluctuation from the environment is usually modeled as 'global' noise. Both intrinsic and extrinsic fluctuations are assumed to be space and time uncorrelated. The strength and properties of fluctuation have a dramatic effect on the evolution of gene network.[27, 28, 29, 30, 31]

The dynamics of the network is usually assumed to be Markovian. That is, at any given time, the future state of the system at the next time step only depends on the current state and doesn't depend on any information of the past or the momentum. The deterministic part of the dynam-

ics can be modeled with Ordinary Differential Equation (ODE).[32, 33]

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) \tag{1.1}$$

x means the protein or RNA (corresponds to each node in the network) concentration. For a N-network that has N nodes, x will be an N-vector. On the left hand side is the shift of the system during the next time step. On the right hand side $F(x)$ is a deterministic force that is a function of x. The dependence on x reflects the nature of bio-chemical reactions involved. Certain reactions are more likely to happen when concentrations of proteins involved are high or low.

This type of determinist equation is widely used in the study of many different system dynamics. It is used in the study of linear and non-linear dynamics[34] and systems like chaos, biological and social networks. The fact that the deterministic force depends only on x reflects the Markovian property of deterministic dynamics. Driving force doesn't know the 'momentum' and has no memory of the past. The 'attractor' of such system can be identified as solutions of $F(x) = 0$. For gene network this corresponds to possible biological states. After long time evolution the network reaches steady state, the stable gene expression levels quantified by x, or the location in the phase space, tends to be at those 'attractors'.

The dynamics of gene network, as mentioned, is stochastic. For this purpose the dynamic equation has to be stochastic instead of deterministic. The Stochastic Differential Equation (SDE) is usually used to model stochastic dynamics:[27, 35]

$$\dot{\mathbf{x}} = F(\mathbf{x}) + \eta(\mathbf{x}, \mathbf{t}) \tag{1.2}$$

This is also known as Langevin equation.[36] In this equation, $\eta(x, t)$ is the intrinsic or extrinsic fluctuation. It is also an N-vector that corre-

sponds to N nodes of the network. When take average over the distribution of fluctuation, we have the fluctuation variance $< \eta(x,t)\eta(x',t') >= D_{ij}(x)\delta_{xx'}\delta_{tt'}$. $D_{ij}(x)$ is an $N \times N$ matrix function of x known as diffusion coefficients/matrix. It is symetric and measures the strength and distribution of the fluctuation. The two delta functions indicate fluctuations are neither time or space correlated.

In the study of gene network, one important task is to identify all possible biological states. In the case of zero or weak fluctuation this can be done by solving ODE $\mathbf{F}(x) = 0$. We can also study the local stability around the 'attractor'. However, when fluctuation is present, the emergence of fluctuation makes the trajectory of the system unpredictable. Even if we prepare two systems at exactly the same initial condition, the evolution results will be very different. The local attractor of ODE doesn't have to be the final stable states.

To study global distribution of biological states, as well as global stability and thermodynamical properties, instead of trajectory which is stochastic and local we can study the evolution of the system from the probability evolution point of view. We will see the landscape theory based on steady state probability distribution provides a global picture of biological states distribution, and unveils several thermodynamical properties of gene network.

1.3 Adiabatic and Non-Adiabatic Dynamics

The stochastic dynamics of gene network includes multiple bio-chemical reactions. It must include synthesis and degradation of transcription factors (TF) and regulatory processes like binding/unbinding of transcription factors to genes. While they all play important roles in shaping epigenetic

landscape,[37] they can be grouped into regulatory processes that change gene states (activate or repress), and protein synthesis/degradation. This introduces a hierarchy in the timescales involved.

In the conventional study, due to the complexity of the system a simpler circumstance is usually considered where the protein translation/degradation is much slower than the regulatory processes. This is called adiabatic limit.[19, 38, 39, 40, 41] Under adiabatic limit, the gene states, or the ensemble of DNA occupancies can be thought to reach quasi-equilibrium as a result of frequent switches. The gene network evolution can be thought as a birth-death problem in protein concentration space with averaged synthesis rate that is determined by the concentration of transcription factors.

The adiabatic limit approximation works well for prokaryotic cells. It greatly simplifies the dynamical equation from size exponential of N master equation to size linear in N Fokker-Planck equation. But for eukaryotic cells, processes like chromatin de-condensation and folding of DNA can reduce the binding/unbinding speed and a lot of times we are at non-adiabatic regime. [42, 43] In order to understand eukaryotic gene network dynamics, we need to consider binding/unbinding processes at least at comparable rates to transcription factor synthesis/degradation.

While some of the non-adiabatic stochastic system dynamics has been studied numerically like electron transition, network dynamics, molecular motors[19, 20, 41, 44, 45, 46], there's not a global theoretical framework to quantify the global properties of biological systems under non-adiabatic dynamics. For non-adiabatic gene network dynamics, adiabatic approximation is no longer valid. The landscape can still be constructed numerically, but the relationship between landscape potential and dynamics is not clear and global properties like steady state flux, global stability and thermodynamical properties like transition can be hard to quantify. On the numerical end the master equation that governs the dynamics grows

exponentially in size and Monte Carlo simulation quickly becomes inefficient.

In the third chapter we develop a theoretical framework by using the similarity between gene network dynamics and quantum mechanics. In the fourth chapter we do a detailed study of non-adiabatic dynamics in a cancer network and unveil its relationship to cancer heterogeneity.

Chapter 2

Dynamics Properties And Calculation Of Transition Rate

2.1 Landscape And Flux Theory

The dynamics of stochastic system can usually be modeled with Langevin equation[34, 47]:

$$\dot{\mathbf{x}} = F(\mathbf{x}) + \eta(\mathbf{x}, \mathbf{t}) \tag{2.1}$$

Langevin equation can be interpreted as the trajectory of stochastic system evolution in phase space. At each time step, system undergoes a deterministic drift governed by driving force $\mathbf{F}$, and a random drift determined by diffusion coefficient.

The trajectory picture clearly demonstrates different roles deterministic force and random fluctuation play in system dynamics. However it is not helpful when we want to study global properties like distribution of stable states, robustness and thermodynamical properties like transition paths and rates. We need a probability description in phase (protein concentration) space. The probability of finding the system being at a specific

point in phase space can be defined as[48]

$$P(x,t) =< \delta(x - x'(t)) >_\eta$$

Here $x'(t)$ is a solution of the dynamical equation and $< \ldots >_\eta$ means average with respect to the noise. We require P to be properly normalized $\int P(x,t)dx = 1$.

In the case there's no fluctuation, the dynamic equation is deterministic without fluctuation component.$\dot{x} = F(x)$. Average over noise in the expression of probability distribution can be removed $P(x,t) = \delta(x - x'(t))$.

The time derivative of P can be written as

$$\begin{aligned}
\partial_t P(x,t) &= -\dot{x'}\frac{d}{dx}\delta(x - x'(t)) \\
&= -F(x')\frac{d}{dx}\delta(x - x'(t)) \\
&= -\frac{d}{dx}(F(x')\delta(x - x'(t))) \\
&= -\frac{\partial}{\partial x}(F(x)P(x,t))
\end{aligned}$$

This describes the probability evolution under a deterministic drifting force. It is a special case of Fokker Planck equation when the fluctuation is turned off. When we have both deterministic force and fluctuation, we need to use stochastic calculus and consider Kramers Moyal expansion of Langevin equation explicitly. Under the same spirit, it can be shown that the Langevin equation that describes trajectory is equivalent to the Fokker-Planck equation that describes probability evolution[49]:

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}(\mathbf{F}(x))P(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\mathbf{D}(\mathbf{x})\mathbf{P}(\mathbf{x},\mathbf{t})) \tag{2.2}$$

The idea of landscape was introduced in the study of protein folding and dynamics.[50] For a quasi-equilibrium system, there exists a potential

energy surface. The evolution of the system is like a charged (gravitaional or electrical) particle moving on the energy surface. The local minimum of the potential energy can be identified as stable states, they are local attractors that traps the system. The system evolves in a gradient way. It always escapes high potential mountains and gets attracted by low potential basins.

Landscape picture is a very helpful method in studying global properties of stochastic biological systems. The topography of the energy surface helps us visualize the distribution of stable biological states in phase space. It helps us identify possible transition paths and intermediate states. Relationship between potential energy function and system dynamics can be constructed so that we know better how dynamics shapes landscape, which is of both scientific and clinical interest.

Without loss of generality, we will take a Fokker-Planck system as an example where there is a clear relationship between landscape potential and dynamics. We will discuss different cases with detailed balance kept/broken and if there is emergence of non-zero steady state flux. The idea of landscape is generalized to gene network. A potential energy surface can be constructed where local basins are identified as possible phenotypic states. As we will see, for gene network in general it is hard to construct relationship between landscape potential and dynamics. It's hard to quantify steady state flux and thermodynamical properties.

For the Fokker Planck system we have above with driving force $\mathbf{F}$ and diffusion matrix $\mathbf{D}$, its physical meaning is how probability distribution $P(x,t)$ evolves with time. On the right hand side of Fokker Planck equation, the first term is the drift term associated with deterministic driving force and the second term is the diffusion term associated with stochastic fluctuation.

Compared with probability conservation equation, the Fokker-Planck

equation can be written in a similar way with right hand side as divergent of a flux term.

$$\frac{\partial}{\partial t}P(x,t) + \nabla \cdot \mathbf{J}(x,t) = 0$$

Here, the flux $\mathbf{J}(x,t)$ is N-dimensional vector and depends on x and t. It is a probability flux that measures the probability flows in and out of the point in phase space.

After long time evolution, the probability distribution will reach steady state $\dot{P}_{SS}0$. The probability conservation equation implies the steady state flux is divergent free $\nabla \cdot \mathbf{J}_{SS} = 0$. For equilibrium system, steady state $\mathbf{J}_{SS}$ itself is zero. There's no flux at steady state, detailed balance is constructed and the system is at equilibrium. A clear analytical relationship can be constructed between landscape potential and dynamics.

For more compact and concise results, we absorb the factor $\frac{1}{2}$ into the definition of $\mathbf{D}$ just for this section. When system reaches steady state with $\dot{P}_{SS} = 0$, the steady state probability distribution P_{SS} doesn't change with time. If at steady state the system is at detailed balance with $\mathbf{J}_{SS} = 0$, we have $\mathbf{F}P_{SS} - \partial(\mathbf{D}P_{SS}) = 0$. In the case $\mathbf{D}$ is constant which means the fluctuation is uniform and global, this indicates the driving force is a pure gradient of a potential function $\mathbf{F} = \mathbf{D}\partial \ln(P_{SS})$. Further we can see the steady state probability distribution has the form of $P_{SS} = e^{\mathbf{D}^{-1}\int \mathbf{F}\cdot d\mathbf{x}}$. We can identify the landscape potential formally as $V = -\int F/D \cdot dx$ and $P_{SS} = e^{-V}$. It is clear from this equation that landscape potential is related to driving force $\mathbf{F}$ as well as fluctuation $\mathbf{D}$. $\mathbf{F}$ plays the role of a conservative force field that can define a potential when integrate over x. $\mathbf{D}$ plays the role similar to a 'metric' in a curved space. For such system, partition function and free energy can be constructed, equilibrium statistical mechanics can be applied and thermodynamical quantities like the global stability, global phase, phase diagram are studied.[23, 51, 52, 53, 54]

When the detailed balance is broken, non-trivial flux emerges as a sign

of system being non-equilibrium. Statistical mechanics for equilibrium system can not be applied and quantifying global thermodynamical quantities can be challenging. First of all, even at steady state when P_{SS} doesn't change with time, steady state flux $\mathbf{J}_{SS}$ is non-zero. The driving force is no longer gradient and will include a curl term. In the case diffusion matrix is constant we have $\mathbf{F} = \mathbf{D}\partial \ln(P_{SS}) - \mathbf{J}_{SS}/P_{SS}$. The driving force decomposition implies that driving force contains a gradient part that plays the role of conservative force field which defines the potential energy. It also contains a curl term that is proportional to the steady state flux. The non-equilibrium potential is related to steady state probability distribution in a similar manner as equilibrium potential does. But this time non-equilibrium potential is related to non-zero steady state curl flux which is not known priori. The stochastic dynamics of biological system is not determined by the gradient potential alone, but also by the non-zero flux. The emergence of steady state flux is a character of system being non-equilibrium and is not present when system is equilibrium with detailed balance.[22, 23, 52, 54, 55, 56, 57]

Gene network is a non-equilibrium stochastic system. Moreover it is not governed by Langevin or Fokker-Planck equations (gene network is governed by chemical master equation). Landscape potential can still be constructed numerically but it can be hard to build a relationship between landscape and dynamics. We can formally define landscape potential as $V = -\ln(P_{SS})$ so that we still have $P_{SS} = e^{-V}$. P_{SS} can be achieved numerically through Monte Carlo simulation[58, 59, 60] so that landscape can be constructed numerically. However, it is not clear in this case how landscape potential is related to driving force and fluctuation. Or in other words, how driving force and fluctuation will affect landscape, stable states, steady state flux, global stability and other thermodynamic quantities.

Equilibrium and non-equilibrium system landscapes are different in many ways. From the construction steps, equilibrium landscape can be constructed directly from underlying driving force and diffusion matrix. Driving force contains only gradient term. Similar to conservative force in physics it can define a potential energy whose local minimum corresponds to a stable biological state of the network. The stochastic dynamics is purely determined by gradient force. Non-equilibrium landscape, on the other hand, is not known priori. It needs to be constructed from steady state probability distribution. The driving force in this case contains conservative gradient force as well as curl flux force. The steady state is non-equilibrium with global non-zero flux that breaks detailed balance. The steady state flux is a measure of frequent matter and energy exchange with outer biological environment. The stochastic dynamics of non-equilibrium system is governed by gradient force and curl flux.

2.2 Transition Rate And Path Integral

An important problem in gene network dynamics is to study transitions between stable biological states induced by noise. In this section we give a brief introduction to the escaping problem and show how it is related to path integral method. We will apply this method in the next two chapters when we study transition paths and rates.

Assume the network we have, when reaches steady state after long time evolution, shows multiple stable states. If we prepare the system at one of the stable states, when the fluctuation is small (temperature is low), There's a hierarchy in timescales. The distribution thermalize inside the well in a short time. But it takes a relative long time to escape the well due to the barrier.

Assume initially the system is at a stable state with protein concentration x_1, the barrier top is x_s. x_s is the boundary of the escaping problem.[61] If given infinite long time, the system will surely escape to infinity.

The escaping rate is defined as:

$$R = \frac{\int_{\text{boundary}} j ds}{\int_{x_1 \text{ well}} \rho dx} \tag{2.3}$$

From quantum mechanics we know the out going current corresponds to the imaginary part of energy.

$$-2ImE \int_{-\infty}^{x_s} |\psi|^2 dx = \frac{1}{2im} (\psi^* \nabla \psi - \psi \nabla \psi^*) = j \tag{2.4}$$

In one dimension this is equivalent to:

$$R(E_n) = j[\int dx |\psi|^2]^{-1} = -2ImE_n \tag{2.5}$$

So escaping rate is proportional to thermal weighted average of the imaginary part of energy, that is the imaginary part of partition function. The problem becomes how to get the imaginary part of partition function.

$$R = -2Im[\sum_n Z^{-1} e^{-E_n/D} E_n] = 2Im[\frac{\partial}{\partial \beta} \ln Z] \tag{2.6}$$

here $\beta = \frac{1}{D}$. Assume we have two stable states x_1 and x_2 at steady state of gene network evolution. In landscape picture, the steady state probability distribution has the form $P_{SS} = e^{-V}$, where x_1 and x_2 are the two local minimum of landscape potential V.

This time initial distribution cannot escape to infinity and energy is real. We prepare the system initially at one stable state x_1. It reaches equilibrium inside x_1 potential well within a short time. Because of the barrier

between x_1 and x_2 is high, it takes a relative long time T for the 'particle' to jump to x_2 potential well. The escaping rate will be defined in the T timescale while the probability distribution mostly remains inside x_1 potential well at equilibrium while there's a small flux penetrates the barrier.

We can calculate transition rate by studying how initial distribution decays with time.[62]

$$\langle x_1, t_f = T| \, e^{-HT} \, |x_0, t_i = 0\rangle$$

For a Fokker-Planck system we discussed in the last section.

$$\partial_t P = -\nabla(FP) + \frac{1}{2}\nabla^2 P \tag{2.7}$$

Define 'momentum operator' $p = -\nabla$, the above equation has the form of Schrodinger's Equation:

$$\partial_t P = [\frac{1}{2}Dp^2 + p \cdot F]P = \hat{H}P \tag{2.8}$$

$\hat{H}$ is the Hamiltonian operator.

This tells us the probability to propagate from one point to another point near it is:

$$\langle x_1| \, 1 + H\delta t \, |x_2\rangle \approx 1 - L\delta t \tag{2.9}$$

where L is Lagrangian:

$$L = \frac{1}{2D}(\dot{x} - F)^2 \tag{2.10}$$

This naive derivation didn't consider stochastic properties of the sys-

tem. To count stochasticity in, instead of normal integral we have to use Ito calculus or Stratonovich integral. It is proved[63] the correct Lagrangian under stochastic calculus is:

$$\mathcal{L} = \frac{1}{2D}(\dot{x} - F)^2 + \frac{1}{2}\nabla \cdot F \tag{2.11}$$

Path integral tells us the probability to propagate from initial point$(x_i, t_i = 0)$ to final point $(x_f, t_f = T)$ is:

$$P = \langle x_f | \, e^{-HT} \, | x_i \rangle = N \int [Dx] e^{-S} \tag{2.12}$$

On the right hand side S is action that follows a specific path:

$$S = \int_0^T dt \mathcal{L}(x, \dot{x}, t) \tag{2.13}$$

In order to calculate the transition amplitude, we have to integrate over all possible paths. Each path contributes a factor that is proportional to the exponential of its negative action. The effective hamiltonian can be constructed follow standard Legendre transformation.

$$\begin{aligned} H &= \frac{1}{2D}\dot{x}^2 - \frac{1}{2D}F^2 - \frac{1}{2}\nabla \cdot F \\ &= \frac{1}{D}\left(\frac{1}{2}\dot{x}^2 - V_{eff}\right) \end{aligned} \tag{2.14}$$

V_{eff} is the effective potential. The effective potential is different from the landscape potential.

$$V_{eff} = \frac{1}{2}F^2 + \frac{D}{2}\nabla \cdot F \tag{2.15}$$

The path integral formalism tells us: a dissipative bio-system under FPE with landscape potential V is equivalent to a charged particle moving in effective potential V_{eff}, at least in the sense of path integral. We also notice that when diffusion matrix $\mathbf{D}$ is small, that is, in the limit where fluctuation is weak. V_{eff} is governed by the first term. The two stable points have $F = 0$, correspond to $V_{eff} = 0$. This agrees with our earlier dissicusion that in the deterministic limit the solution of ODE corresponds to stable states. The saddle point of the barrier also has $F = 0$, thus $V_{eff} = 0$. The effective potential has the shape of 'W'.

Now we can calculate the 'stay' probability. The largest contribution comes from oscillations around the stable point, and this part is trivial.

The next leading contribution comes from the particle initially at x_1, going to the saddle point x_s and bounce back. We use instanton/anti-instanton method to calculate this.

Instanton is the classical solution of equation of motion, connecting initial point x_1, t_i and final point x_2, t_f. We get this solution by minimizing action. We will call instanton trajectory $x_c(t)$ and corresponding action S_0.[64]

According to Caroli[65], we only need to consider instanton of infinite large time $T \to \infty$, or instanton with zero energy for our own purpose. We use semi-classical approximation, that is only consider contribution from trajectories around instanton:

$$x(t) = x(t)_c + y(t) \tag{2.16}$$

In general y can be expanded in eigenmodes:

$$y = \sum_n \lambda_n y_n \tag{2.17}$$

$$y_n(t=0) = y_n(t=T) = 0 \tag{2.18}$$

$$\int_0^T y_n(t) y_m(t) = \delta_{n,m} \tag{2.19}$$

Use that $\frac{\delta S}{\delta x}|_{x=x_c} = 0$, and keep the second order of y in the exponential, the probability becomes:

$$\begin{aligned} P = \quad & Ne^{-S_0} \int [Dy] \exp -\{\frac{1}{D}\int_0^T dt(\dot{y}^2 - (\nabla \cdot F)\dot{y}y \\ & + \frac{1}{2}\nabla^2(\frac{1}{2}F^2 + \frac{D}{2}\nabla \cdot F)y^2)\} \end{aligned} \tag{2.20}$$

the second term in the exponential is zero as:

$$\int_0^T dt(\nabla \cdot F(x_c))\dot{y}y = \frac{1}{2}(\nabla \cdot F(x_c))y^2\Big|_{t=0}^{t=T} = 0 \tag{2.21}$$

The survival probability becomes:

$$\begin{aligned} P &= Ne^{-S_0} \int [dy] \exp - \int_0^T dt[\frac{1}{2D}(\dot{y}^2 - V''y^2)] \\ &= Ne^{-S_0} \det[\frac{1}{D}(-\partial_t^2 + V'')]^{-1/2} \end{aligned} \tag{2.22}$$

This is the contribution from one instanton. Similarly there will be anti-instanton going from x_f to x_i. Because of the emergence of flux, instanton and anti-instanton have different trajectories. Contribution from one anti-instanton can be written as:

$$Ne^{-S_0^*} \det[\frac{1}{D}(-\partial_t^2 + V'')^*]^{-1/2} \tag{2.23}$$

Instantons and Anti-instantons are well localized pseudo-particles in time axes. To ensure we get to x_f at the end, the trajectory must be $x_i \to x_f \to x_i \to x_f \to x_i \ldots \to x_f$ or the distribution of instanton, anti-instanton must be $i \to a \to i \to a \ldots \to i$.

Integrate over the locations of centers gives us:

$$\int_0^T dt_1 \int_0^T dt_2 \ldots \int_0^T dt_n = \frac{T^n}{n} \tag{2.24}$$

In the case of multi instanton we have to consider corrections from finite transition time of each transition. Contribution from instanton and anti-instanton becomes:

$$\begin{aligned} P_i &= NKe^{-S_0} \\ P_a &= NK^*e^{-S_0^*} \end{aligned} \tag{2.25}$$

Sum over all such trajectories, we get the survival probability as well as transition rate R.

$$\begin{aligned} P &= N \sum_{evenn} \frac{(\sqrt{KK^*}e^{-\frac{1}{2}(S_0+S_0^*)}T)^n}{n!} \\ &= N(e^{RT} + e^{-RT}) \end{aligned} \tag{2.26}$$

2.3 Calculation Of Transition Rate

We first explain why we only consider instanton with zero energy. Strictly speaking, to calculate transition rate we have to calculate:

$$\int_0^\infty dT e^{-S(T)} \tag{2.27}$$

No matter how large the exponential is when T is finite, when the integral goes to infinity, it converges to:

$$\int_0^\infty dT e^{-S(T)} \to T e^{-S(\infty)} \tag{2.28}$$

which means the integral is controlled by $S(\infty)$.

We have to calculate two things: the prefactor N and the exponential factor $K(K^*)$.

Calculation of N is somewhat standard. Assume effective potential V can be approximated as $V \approx \frac{1}{2}\omega r^2$. For large T, we have:

$$N[det(-\partial_t^2 + \omega^2)]^{-1/2} = (\frac{\omega}{\pi})^{1/2} e^{-\omega T} \tag{2.29}$$

On the other hand, by carefully integrate over zero mode and compare with one instanton case we can get:

$$K = (\frac{S_0}{2\pi})^{1/2} \left| \frac{det(-\partial_t^2 + \omega^2)}{det'(-\partial_t^2 + V'')} \right|^{1/2} \tag{2.30}$$

$$K^* = (\frac{S_0^*}{2\pi})^{1/2} \left| \frac{det(-\partial_t^2 + \omega^2)}{det'(-\partial_t^2 + V'')^*} \right|^{1/2} \tag{2.31}$$

According to Coleman[64], if we can construct solution of equation:

$$(-\partial_t^2 + W)\psi = \lambda\psi \tag{2.32}$$

with boundary condition:

$$\psi_\lambda(t=0) = 0 \tag{2.33}$$

$$\partial_t\psi_\lambda(t=0) = 1 \tag{2.34}$$

We have:

$$\frac{det(-\partial_t^2 + W^1)}{det(-\partial_t^2 + W^2)} = \frac{\psi^1(T)}{\psi^2(T)} \tag{2.35}$$

thus:

$$\frac{det(-\partial_t^2 + \omega^2)}{\psi_0(T)} = \pi N^2 \tag{2.36}$$

By approximate effective potential as harmonic oscillator around stable point and construct wave function ψ_0 we see we get correct normalization constant N.

Similarly for the other determinant, we want to construct solution from two zero modes of fluctuation equation. When T is large, zero mode behaves as:

$$x_1 = S_0^{-1} d\bar{x}/dt \to Ae^{-|t|} \tag{2.37}$$

$$y_1 \to Ae^{\pm|t|} \tag{2.38}$$

A can be determined from 'time translation' zero mode:

$$t = \int_0^{\bar{x}} dx(2V)^{-1/2} = -\ln[S_0^{-1/2} A^{-1}\dot{\bar{x}}] \tag{2.39}$$

Solution of smallest eigenvalue reads:

$$\psi_0 = (2A)^{-1}(e^{T/2}x_1 + e^{-T/2}y_1) \tag{2.40}$$

And finally:

$$\frac{det'(-\partial_t^2 + V'')}{det(-\partial_t^2 + \omega^2)} = \frac{\psi_0(T)}{\lambda_0 e^{T/2}} = \frac{1}{2A^2} \tag{2.41}$$

This method provides a way to calculate prefactor. Next we discuss how to calculate S_0 which is the action of classical path. The effective action $S = \int Ldt$ is the exponential part of the path integral which calculates the probability of the system travels from initial point to final point. In order for the path to make dominant contribution among all paths, the corresponding action should be minimized. From the effective Lagrangian we can get conjugate momentum to be:

$$p = \frac{\partial}{\partial x}L = \frac{1}{D}(\dot{x} - F)$$

The hamiltonian is

$$E = p\dot{x} - L = \frac{1}{2D}\dot{x}^2 = \frac{1}{2D}\dot{x}^2 - \frac{1}{2D}F^2 - \frac{1}{2}D\partial(F/D) = \frac{1}{2D}\dot{x}^2 - V_{eff}$$

We can see the inverse of diffusion matrix $\mathbf{D}^{-1}$ plays the role of 'metric' in a curved space. We can intorduce line element in curved space as $dl = \sqrt{(\frac{1}{D})_{\mu\nu}dx_\mu dx_\nu}$. There are two terms on the right hand side. The first term can be explained as kinetic energy term. The second depends only on x and can be interpreted as an effective potential energy. The energy is

conserved for classical path. From above equation we get

$$dt = \sqrt{\frac{D^{-1}_{\mu\nu} dx_\mu dx_\nu}{2(E + V_{eff})}} = dl \sqrt{\frac{1}{2(E + V_{eff})}} \tag{2.42}$$

We are considering action of classical path. The integral is challenging since we are integrating over infinite long time and infinite slow speed. Luckily we can transform the integral over time into line integral in curved space. Use the expression of the line element in curved space and dt, the action can be transformed into integral in space[66, 67] :

$$S = \int p\dot{x}dt = \int_{x_i}^{x_f} \sqrt{2(E + V_{eff})} dl - \int_{x_i}^{x_f} F/D dx \tag{2.43}$$

The path that minimizes above action corresponds to the dominant path connecting initial and final states. The calculation of prefactor and dominant path action complete the calculation of the transition rate. We will apply this method in the next two chapters.

Chapter 3

Multiple coupled landscapes and non-adiabatic dynamics

3.1 Introduction

In physical and biological systems that have frequent energy and matter exchange with environments (e.g. gene regulatory networks), stochasticity is often unavoidable with intrinsic and extrinsic fluctuations [68, 69]. These stochastic processes have been modeled mathematically as drifted Brownian motions, where the system evolves in a way similar to a point particle moving diffusively along an underlying landscape. This similarity has led to a landscape picture where steady states are identified as local minimums and non-equilibrium dynamics as flows on the landscape determined by the landscape gradient and a curl flux measuring the degree of deviation from detailed balance [20, 23, 24, 53, 70].

A single landscape, however, is not sufficient for describing the process which is dynamically modulated by the discrete change in the system

state as in gene regulatory networks [39, 71, 72] and molecular motors [73, 74, 75]. By regarding discrete change as hopping between different landscapes, these processes are described by coupled multiple landscapes, which are characterized with different timescales, timescale within each landscape and timescale of transitions among landscapes.

In order to quantify this hierarchy in timescales, we introduce an adiabaticity parameter ω. Large ω corresponds to the adiabatic limit where the inter-landscape hoping is more frequent than the intra-landscape motion. Due to the rapid transitions, a single effective landscape emerges from the 'average' of multiple landscapes. In the small ω non-adiabatic limit, the description can be simplified as the system traps in one of the discrete landscapes for a long time [19, 20, 39, 41, 71, 72]. In the moderate non-adiabatic region, however, a proper physical understanding and analytical treatment is still challenging as numerical simulation quickly becomes inefficient as system size grows.

In this study, we use a N-gene regulatory network as an example to show that by developing a continuous spinor representation and the associated path integral method, the underlying 2^N discretely (e.g. each gene has 2 discrete states: on and off) coupled continuous stochastic processes can be mapped onto stochastic continuous process in a $2N$ dimensional extended space. In other words, there exists mathematical mapping of the dynamics on 2^N discretely coupled landscapes onto the one single landscape in $2N$ dimensional extended space. Our work, for the first time, provides a general analytic treatment of non-adiabatic dynamics for the discretely coupled continuous non-equilibrium stochastic systems. It gives a physical and quantitative picture to understand dynamics in non-adiabatic regime. On this $2N$ dimensional landscape, eddy current emerges as a sign of non-equilibrium non-adiabatic dynamics and plays an important role in system evolution. We also generalize our re-

sult to the general multi-gene network with each gene of having multiple binding sites or multiple activated states, so that it can be applied to variety of gene networks and discretely coupled continuous non-equilibrium stochastic physical and biological systems.

Many interesting physical effects emerge due to non-adiabaticity manifested by the eddy current in extended space. We do a detailed numerical study of $N = 1$ self-activator as a special case of our general treatment. Our theory is able to explain the enhancement of fluctuations in non-adiabatic region. Most importantly by extending to 2-dimensional space, the transition between 'on' and 'off' states becomes irreversible due to flux emergence in 2-dimensional space. There is energy or heat dissipation measured by the entropy production associated with this irreversibility.[23] An optimal transition rate can appear at weak non-adiabatic region. We can see different behavior for 'on' to 'off' and 'off' to 'on' transitions due to irreversibility. The optimal rate can play an important role in gene network evolution and similar biological systems.

3.2 Model

To demonstrate our point, we study an N-gene regulatory network. Other physical and biological systems work in a similar way (e.g. N-electronic surfaces). For this specific network, each gene has one binding site on which one of the transcription factors can be bound. As a result, each gene has discrete on/off states that have different transcription expression levels [76]. The whole N-gene network has 2^N discrete states. We use s to denote a specific discrete (gene on and off) state of the network, and $P_s(n,t)$ to denote the probability of the system at state s with protein numbers $n = (n_1, n_2, \ldots n_N)$. The master equation is 2^N dimensional. Especially for

$N = 1$ self activator motif where the transcription factor is a dimer, the master equation is

$$\begin{aligned}
\partial_t P_1(n) &= \frac{h_0}{2}(n+2)(n+1)P_0(n+2) - fP_1(n) \\
&+ g_1(P_1(n-1) - P_1(n)) + k((n+1)P_1(n+1) - nP_1(n)) \\
\partial_t P_0(n) &= -\frac{h_0}{2}n(n-1)P_0(n) + fP_1(n-2) \\
&+ g_0(P_0(n-1) - P_0(n)) + k((n+1)P_0(n+1) - nP_0(n))
\end{aligned} \tag{3.1}$$

Here g/k is the protein synthesis/degradation rate, h/f is the binding/unbinding rate. At large volume limit, protein concentration $x = \frac{n}{V_0}$ becomes continuous. After dropping out higher order terms of $\frac{1}{V_0}$ master equation becomes:

$$\begin{aligned}
\partial_t \begin{pmatrix} P_1(n) \\ P_0(n) \end{pmatrix} &= \begin{pmatrix} -\partial_x F_1 + \partial_x^2 D_1 & \\ & -\partial_x F_0 + \partial_x^2 D_0 \end{pmatrix} \begin{pmatrix} P_1(n) \\ P_0(n) \end{pmatrix} \\
&+ \begin{pmatrix} -f & h \\ f & -h \end{pmatrix} \begin{pmatrix} P_1(n) \\ P_0(n) \end{pmatrix}
\end{aligned} \tag{3.2}$$

Where we define $h = \frac{h_0}{2}n^2$, redefine $k = k \cdot V_0$. The driving forces and diffusion coefficients are $F_1 = \frac{1}{V_0}(g_1 - kx)$, $F_0 = \frac{1}{V_0}(g_0 - kx)$ and $D_1 = \frac{1}{2V_0^2}(g_1 + kx)$, $D_0 = \frac{1}{2V_0^2}(g_0 + kx)$.

This equation has the form of coupled Fokker-Planck equation. The first operator matrix $\mathcal{H}_0$ defines 2 discrete Fokker-Planck landscapes. The second operator matrix $\mathcal{H}_b$ describes the hopping processes that couple the discrete landscapes. As we can see from Fig. 1, at $\omega = 0.001$ when the coupling is weak, steady state shows two Fokker-Planck basins. At $\omega = 1000$ when the coupling is strong, the two basins merge into one.

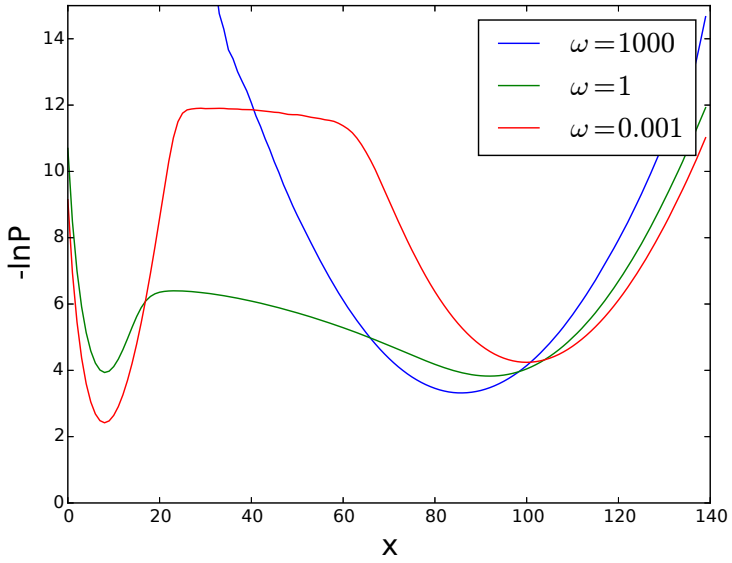


Figure 1: One-dimensional landscape of self activator at different adiabaticity

In a more general N-gene network, the dynamic master equation is 2^N dimensional. For each of the s-components which corresponds to a specific

spin state we have:

$$
\begin{aligned}
\frac{\partial}{\partial t}P_s(n,t) \quad &= \sum_{i \text{ for } s_i=1} h_{ij}P_{s'}(n_1,\dots n_{i-1},n_i-2,n_{i+1},\dots n_N,t) \\
&+ \sum_{i \text{ for } s_i=0} f_{ij}P_{s''}(n_1,\dots n_{i-1},n_i+2,n_{i+1},\dots n_N,t) \\
&- \sum_{i \text{ for } s_i=0} h_{ij}P_s(n_1,\dots n_{i-1},n_i,n_{i+1},\dots n_N,t) \\
&+ \sum_{i \text{ for } s_i=1} f_{ij}P_s(n_1,\dots n_{i-1},n_i,n_{i+1},\dots n_N,t) \\
&+ \sum_{i \text{ for } s_i=1} g_1^i(P_s(n_1,\dots n_{i-1},n_i-1,n_{i+1},\dots n_N,t)-P_s(n,t)) \\
&+ \sum_{i \text{ for } s_i=0} g_0^i(P_s(n_1,\dots n_{i-1},n_i-1,n_{i+1},\dots n_N,t)-P_s(n,t)) \\
&+ k^i((n_i+1)P_s(n_1,\dots n_{i-1},n_i+1,n_{i+1},\dots n_N,t)-n_iP_s(n,t))
\end{aligned}
\tag{3.3}
$$

Here s' corresponds to the spin configuration where $s'_i = 0$ while $s_i = 1$ and the rest components of s and s' are the same. s'' corresponds to the spin configuration where $s''_i = 1$ while $s_i = 0$ and the rest equal. In the case the transcription factor is dimer h_{ij} has the form of $\frac{\tilde{h}_{ij}}{2}n_j(n_j-1)$. In the large volume limit, protein concentration $x_i = \frac{n_i}{V_0}$ becomes continuous. The master equation above has the form of coupled Fokker-Planck equation:

$$
\begin{aligned}
\frac{\partial}{\partial_t}P_s(x,t) \quad &= \sum_{i \text{ for } s_i=1} h_{ij}P_{s'}(x,t) + \sum_{i \text{ for } s_i=0} f_{ij}P_{s''}(x,t) \\
- \quad & \sum_{i \text{ for } s_i=0} h_{ij}P_s(x,t) - \sum_{i \text{ for } s_i=1} f_{ij}P_s(x,t) \\
- \quad & \frac{1}{V_0}\sum_{i \text{ for } s_i=1} \partial_{x_i}(g_1^i P_s(x,t)) + \sum_{i \text{ for } s_i=1} \frac{1}{2V_0^2}\partial_{x_i}^2(g_1^i P_s(x,t)) \\
- \quad & \frac{1}{V_0}\sum_{i \text{ for } s_i=0} \partial_{x_i}(g_0^i P_s(x,t)) + \sum_{i \text{ for } s_i=0} \frac{1}{2V_0^2}\partial_{x_i}^2(g_0^i P_s(x,t)) \\
+ \quad & \sum_i \frac{1}{V_0}\partial_{x_i}(k^i x_i P_s(x,t)) + \frac{1}{2V_0^2}\sum_i \partial_{x_i}^2(k^i x_i P_s(x,t))
\end{aligned}
\tag{3.4}
$$

Again it has the form of coupled Fokker-Planck equation.

$$
\frac{d}{dt}P_s(n,t) = \sum_{s'} \mathcal{H}_{ss'}P_{s'}(n,t) = \sum_{s'}(\mathcal{H}_{0ss'} + \mathcal{H}_{bss'})P_{s'}(n,t) \tag{3.5}
$$

$\mathcal{H}$ in the above equation can be decomposed as $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_b$, where $\mathcal{H}_0$ is diagonal and describes protein synthesis/degradation, and $\mathcal{H}_b$ represents the binding/unbinding processes with non-diagonal elements describing hopping between discrete on and off gene states. $\mathcal{H}_0$ has terms of g_0^i, g_1^i and k_i, where g_0^i and g_1^i are the protein synthesis rates of i-protein when i-gene is at unbound and bound states, respectively. k_i is the degradation rate of i-protein. In $\mathcal{H}_b$, we use h_{ij} and f_{ij}, where h_{ij} is the binding rate of j-protein with i-gene, f_{ij} is the corresponding unbinding rate. In the case where the transcription factor is dimer, $h_{ij} = \frac{1}{2}\tilde{h}_{ij}n_j(n_j-1)$. f_{ij} eaquals $k \times \omega$ is constant. As mentioned, the key character of such system is the hierarchy of two timescales, one for protein synthesis/degradation and the other for gene activation/repression. To quantify this timescale hierarchy, we introduce adiabaticity parameter as $\omega = f/k$. Other param-

eters we use to set up the networks are $X_{eq} = f/h$ (equilibrium constant of binding), $X_{ad} = (g_0 + g_1)/2k$ (synthesis relative to degradation) and $\delta X = (g_1 - g_0)/2k$ (difference in on and off synthesis). We can make discrete protein copy numbers n_i continuous by considering large volume limit, and replace n_i with concentration $x_i = n_i/V_0$. In this way, we always work in the continuous x-space.

In the self activator example we considered, $\mathcal{H}_0$ defines 2 discrete Fokker-Planck states and $\mathcal{H}_b$ describies the coupling between these states. Adiabaticity ω describes the relative strength of $\mathcal{H}_b$ and $\mathcal{H}_0$. When ω is large, we arrive at the adiabatic limit. From the master equation, $\mathcal{H}_b$ dominates over $\mathcal{H}_0$, so that we are always close to the steady state $\mathbf{r}_0$ of $\mathcal{H}_b$: $\mathcal{H}_b\mathbf{r}_0 = 0$. For self-activator, $\mathbf{r}_0 = (\frac{h}{\sqrt{h^2+f^2}}, \frac{f}{\sqrt{h^2+f^2}})^T$. The master equation becomes $\partial_t P(n,t) = 1^T H_0(P(n,t)\mathbf{r}_0)$, which has the form of one-dimensional FPE [39, 77]. The other limit, when ω is small, corresponds to the weakly coupling case where hopping between different states is rare. $\mathcal{H}_0$ dominates over $\mathcal{H}_b$. Because $\mathcal{H}_0$ is 2 dimensional and diagonal, if we prepare the initial state at one of the 2 discrete states, the system tends to be trapped there for a long time. This is exactly what we see from Fig. 1.

The moderate ω non-adiabatic regime is of particular interest. As has been pointed out, dynamics in this region can be crucial to physical and biological system dynamics [20, 41, 72]. However, in this regime, terms in $\mathcal{H}_0$ and $\mathcal{H}_b$ are comparable, the adiabatic approximation no longer works. For general N-gene network, although the master equation is exact, its size grows as 2^N, numerical simulation quickly becomes non-efficient. A clear physical and analytical picture is needed to find a way around this.

3.2.1 Path Integral and Effective Lagrangian

The master equation has the form of Schrödinger equation. The $2^N \times 1$ state vector $\mathbf{P}$ plays the role of wave function. $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_b$ becomes

the hamiltonian operator when acting on $\mathbf{P}$. Though $\mathcal{H}$ is non-hermitian, reflecting the non-equilibrium features of the system, the change of gene on/off states is similar to the change of spin-up/down states of electron [72]. To quantify the 'spin state' of the ith gene, we use spinor $|s_i\rangle$ parameterised by θ_i and ϕ_i.

$$|s_i\rangle = \begin{pmatrix} \cos^2 \frac{\theta_i}{2} e^{i\phi_i/2} \\ \sin^2 \frac{\theta_i}{2} e^{-i\phi_i/2} \end{pmatrix} \qquad \langle s_i| = \begin{pmatrix} e^{-i\phi_i/2}, & e^{i\phi_i/2} \end{pmatrix} \tag{3.6}$$

With spin up representing the bound state and spin down the unbound state. This spinor is properly normalized: $\langle s^i|s^i\rangle = 1$, the identity operator in spin space can be written as:

$$I_s = \prod_i \frac{1}{4\pi} \int d\cos\theta_i d\phi_i \, |s_i\rangle \langle s_i| \tag{3.7}$$

The amplitude in the two components of $|s_i\rangle$ has the physical meaning of the probability of the corresponding binding site being at bound/unbound state. A general state of the system $|x, s, t\rangle$can be described with protein copy number $x = (x_1, x_2, \ldots x_N)$ and 'spin' state $s = s_1 \otimes s_2 \ldots s_N$.

Path integral method tells us that the transition probability from an initial point with protein concentration $x_i = (x_i^1, x_i^2, \ldots x_i^N)$(with $1, 2, \ldots, N$ labeling genes or proteins) and spin states s_i at time t_i to the final point with (x_f, s_f, t_f) can be calculated by sum over all paths connecting these two points with exponential weight proportional to the integral of effec-

tive Lagrangian along the path [78]. That is:

$$
\begin{aligned}
&P(x_f, s_f, t_f | x_i, s_i, t_i) = \langle x_f, s_f | \, e^{\int_{t_i}^{t_f} \mathcal{H} d\tau} \, | x_i, s_i \rangle \\
&= \langle x_f, s_f | \left[\lim_{\Delta t \to 0} \prod_{\Delta t} I_x \otimes I_s (1 + \mathcal{H} \Delta t) I_x \otimes I_s \right] | x_i, s_i \rangle \\
&= \langle x_f, s_f | \lim_{\Delta t \to 0} \left(\prod_{\tau = t_i + \Delta t}^{\tau = t_f} | p, s, \tau \rangle \langle p, s, \tau | \, (1 + \mathcal{H} \Delta t) \, | p, s, \tau - \Delta t \rangle \langle p, s, \tau - \Delta t | \right) | x_i, s_i \rangle \\
&= \text{const} \prod_i \mathcal{D}x \mathcal{D}p \mathcal{D} \cos\theta_i \mathcal{D}\phi_i e^{-\int \mathcal{L} dt}
\end{aligned}
\tag{3.8}
$$

Here, $\mathcal{L}$ is the effective Lagrangian, which contains $2N$ coordinate q-like variables x_i and c_i with c_i being the probability of the $s_i = 0$ state, and $2N$ momentum p-like variables p_i and ϕ_i.

To simplify our notation we will note $\sin^2(\theta_i/2)$ as c_i, thus $\cos^2(\theta_i/2) = 1 - c_i$ The matrix element $\langle p, s, \tau | \, (1 + \mathcal{H} \Delta t) \, | p, s, \tau - \Delta t \rangle$ receives contribution from 3 parts:

$$
\langle p, s, \tau | \, 1 \, | p, s, \tau - \Delta t \rangle = 1 + i \sum_i p_i \dot{x}_i \Delta t - i \sum_i \phi_i \dot{c}_i \Delta t \tag{3.9}
$$

$$
\begin{aligned}
\langle p, s, \tau | \, \mathcal{H}_0 \, | p, s, \tau - \Delta t \rangle = \quad & \sum_i \frac{1}{V_0} (-i p_i)(g_0^i - k_i x_i) c_i + \sum_i \frac{1}{V_0} (-i p_i)(g_1^i - k_i x_i)(1 - c_i) \\
& - \sum_i \frac{1}{2V_0^2} p_i^2 (g_0^i + k_i x_i) c_i - \sum_i \frac{1}{2V_0^2} p_i^2 (g_1^i + k_i x_i)(1 - c_i)
\end{aligned}
\tag{3.10}
$$

$$
\langle p, s, \tau | \, \mathcal{H}_i \, | p, s, \tau - \Delta t \rangle = \sum_i \sum_j \left(-h_{ij} c_i + h_{ij} c_i e^{-i\phi_i} + f_{ij}(1 - c_i) e^{i\phi_i} - f_{ij}(1 - c_i) \right) \tag{3.11}
$$

The effective Lagrangian $\mathcal{L}$, turns out to be:

$$\begin{aligned} -\mathcal{L} \quad &= \sum_i i p_i \frac{d}{dt} x_i - \sum_i i\phi_i \frac{d}{dt} c_i \\ &\sum_i \frac{1}{V_0}(-ip_i)(g_0^i - k_i x_i)c_i + \sum_i \frac{1}{V_0}(-ip_i)(g_1^i - k_i x_i)(1-c_i) \\ &\sum_i \frac{1}{2V_0^2} p_i^2 (g_0^i + k_i x_i)c_i - \sum_i \frac{1}{2V_0^2} p_i^2 (g_1^i + k_i x_i)(1-c_i) \\ &\sum_i \sum_j \left(-h_{ij}c_i + h_{ij}c_i e^{-i\phi_i} + f_{ij}(1-c_i)e^{i\phi_i} - f_{ij}(1-c_i)\right) \end{aligned} \tag{3.12}$$

Effective Langrangian is equivalent to Hamiltonian in providing dynamics information. Under spinor representation it is possible to map the original dynamics in continuous x space with discrete gene states into continuous dynamics in x space and spin space.

3.2.2 Deterministic Dynamics and Intrinsic Noise

Effective Lagrangian provides information of dynamics in continuous x space as well as in spin space. To reduce the dimensionality and concentrate on the dynamics in the observable space, we can integrate out the conjugate variables. We do so by expanding over conjugate variables p_i and ϕ_i. The first order terms correspond to classic dynamics, or deterministic dynamics:

$$\begin{aligned} -\mathcal{L}_{cl} = \quad &\sum_i i p_i \frac{d}{dt} x_i - \sum_i i\phi_i \frac{d}{dt} c_i \\ &+ \sum_i \frac{1}{V_0}(-ip_i)(g_0^i - k_i x_i)c_i + \sum_i \frac{1}{V_0}(-ip_i)(g_1^i - k_i x_i)(1-c_i) \\ &+ \sum_i \sum_j i\phi_i \left(f_{ij}(1-c_i) - h_{ij}c_i\right) \end{aligned} \tag{3.13}$$

When integrating over ϕ_i and p_i, $\mathcal{L}_{cl}$ contributes $2N$ delta functions, providing $2N$ deterministic equations for x_i and c_i.

$$\dot{x}_i = \frac{1}{V_0}\left(g_0^i c_i + g_1^i(1-c_i) - k_i x_i\right) \quad \dot{c}_i = \sum_j \left(f_{ij}(1-c_i) - h_{ij}c_i\right) \tag{3.14}$$

We notice that though we didn't use self-consistent assumption, at deterministic level we obtain the same result as the self-consistent approach with zero-fluctuation hamiltonian equations [79]: $\dot{x} = \frac{\partial \mathcal{H}}{\partial p}\big|_{p=0}$ and $\dot{c} = \frac{\partial \mathcal{H}}{\partial \phi}\big|_{\phi=0}$.

As we go to the next leading order, we obtain quadratic terms of conjugate variables p_i and ϕ_i. This corresponds to semi-classical dynamics.

$$\begin{aligned} -\mathcal{L}_{sc} = \quad & -\sum_i \frac{1}{2V_0^2} p_i^2 (g_0^i + k_i x_i) c_i - \sum_i \frac{1}{2V_0^2} p_i^2 (g_1^i + k_i x_i)(1-c_i) \\ & -\sum_i \frac{1}{2}\phi_i^2 \left(\sum_j \left(f_{ij}(1-c_i) + h_{ij}c_i\right)\right) \end{aligned} \tag{3.15}$$

After the Hubbard-Stratonovich transformation, quadratic terms provide gaussian intrinsic fluctuations [66]. This suggests the effective dynamics is governed from the original $2^N \times N$ dynamics to now by a $2N$ dimensional coupled Langevin system with coordinate dependent diffusion coefficients:

$$\dot{x}_i = \frac{1}{V_0}(g_0^i c_i + g_1^i(1-c_i) - k_i x_i) + \eta_{xi}, \tag{3.16}$$

$$\dot{c}_i = \sum_{j=1}^{N} f_{ij}(1-c_i) - \sum_{j=1}^{N} h_{ij}c_i + \eta_{ci}, \tag{3.17}$$

with η_{xi} and η_{ci} being intrinsic gaussian noises;

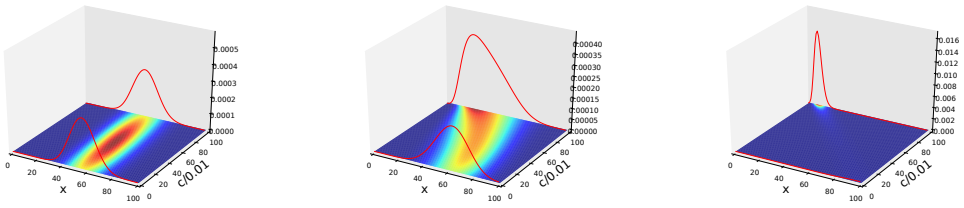
$$\langle \eta_{x_i}(t)\eta_{x_j}(t')\rangle = \frac{1}{{V_0}^2}\left(g_0^i c_i + g_1^i(1-c_i) + k_i x_i\right)\delta_{ij}\delta(t-t'),$$
$$\langle \eta_{c_i}(t)\eta_{c_j}(t')\rangle = \left(\sum_j \left(f_{ij}(1-c_i) + h_{ij}c_i\right)\right)\delta_{ij}\delta(t-t').$$

If we view the problem in terms of probability distribution, the coupled Langevin equations are equivalent to a $2N$ dimensional FPE governed system:

$$\begin{aligned}\partial_t P(n,s,t) = &-\sum_{x_i}\partial_{x_i}(F_{x_i}P) - \sum_{c_i}\partial_{c_i}(F_{c_i}P)\\ &+\sum_{x_i}\partial^2_{x_i}(D_{x_i x_i}P) + \sum_{c_i}\partial^2_{c_i}(D_{c_i c_i}P).\end{aligned} \tag{3.18}$$

Mapping the system stochastic dynamics to the probability evolution via FPE provides a landscape picture for dynamics especially in the non-adiabatic region. Define potential U from the steady state probability distribution ρ^{SS} as $U = -\log\rho^{SS}$. The stable steady states correspond to local minimums of U. The stability of the steady state is determined by the geometry or depth of the basin. Probability conservation tells us the steady state flux is: $j_i^{SS} = -F_i\rho^{ss} + \partial_j(D_{ij}\rho^{ss})$ and $\nabla\cdot\mathbf{j} = 0$. The steady state flux is divergent free and therefore a rotational curl. If we define $\tilde{F}_i = F_i - (\partial_j D_{ij})$, the steady state flux can be written in a way similar to what we have in constant diffusion coefficient case: $\mathbf{j} = -\mathbf{F}\rho^{SS} + \partial\cdot(\mathbf{D}\rho^{SS}) = -\tilde{\mathbf{F}}\rho^{SS} + \mathbf{D}\cdot\partial\rho^{SS}$. The driving force can be decomposed into a gradient part plus a curl part: $F_i = -D_{ij}\partial_j U + j_i^{ss}/\rho^{ss} + \partial_j D_{ij}$ or $\tilde{F}_i = -D_{ij}\partial_j U + j_i^{ss}/\rho^{ss}$. [23]. The system undergoes a drifted Brownian motion on this $2N$-dimensional extended space landscape. If there's only conservative gradient force, the system would be equilibrated at a steady state and there is no flux. The

non-zero steady state flux, or eddy current gives a quantitative measure of the detailed balance breaking. As we sill see in the self activator example, it plays an important role in entropy production and non reversible system dynamics.



(a) $\omega = 1000$ (b) $\omega = 1$ (c) $\omega = 0.01$

Figure 2: Steady state distribution of self-activator in extended space at (a): $\omega = 1000$, (b):$\omega = 1$, (c):$\omega = 0.01$

3.2.3 Generalization to N-gene network with multiple binding sites

The example we considered with each gene has only one binding site is rather simplified. If we consider each gene can have up to M binding sites, then each gene has 2^M discrete spin states. For instance, when $M = 2$, the i-gene state has $2^M = 4$ discrete states: $(P_{s_i=00}(n,t), P_{s_i=10}(n,t), P_{s_i=01}(n,t), P_{s_i=11}(n,t))^T$. These 2^M discrete states can have different protein synthesis rate; instead of g_0^i and g_1^i now we have 2^M different $g_{s_i}^i$. We introduce an extra index $1 \leq \mu \leq M$ to distinguish different binding sites on the same gene. The "spin" state of ith gene can be represented as $|s_i\rangle = \prod_{\mu=1}^{M} |s_i^\mu\rangle$. The identity operator in spin space is:

$$I_s = \prod_{i=1}^{N} \left(\frac{1}{(4\pi)^M} \prod_{\mu=1}^{M} \left(\int_0^1 dc_i^\mu \int_{-2\pi}^{2\pi} d\phi_i^\mu \right) |s_i\rangle \langle s_i| \right)$$

The probability of propagation can still be calculated using path inte-

gral when we pay special attention to the index μ. Follow a similar routine we get effective Lagrangian:

$$\begin{aligned}
-\mathcal{L} = \quad & \sum_i i p_i \dot{x}_i - \sum_i i\phi_i^\mu \dot{c}_i^\mu \\
& + \sum_i \frac{1}{V_0}\left[\sum_{s_i} g_{s_i}^i \left(\prod_{s_i^\mu=1}(1-c_i^\mu)\prod_{s_i^\nu=0} c_i^\nu\right) - k_i x_i\right] \\
& - \frac{1}{2V_0^2}\sum_i p_i^2 \left[\sum_{s_i} g_{s_i}^i \left(\prod_{s_i^\mu=1}(1-c_i^\mu)\prod_{s_i^\nu=0} c_i^\nu\right) + k_i x_i\right] \\
& + \sum_{i,j}\sum_\mu [-h_{ij}^\mu c_i^\mu + h_{ij}^\mu c_i^\mu e^{-i\phi_i^\mu} \\
& + f_{ij}^\mu (1-c_i^\mu) e^{i\phi_i^\mu} - f_{ij}(1-c_i)]
\end{aligned}$$

When expand the effective lagrangian over p_i and ϕ_i^μ up to the second order, we see the coupled FPE system is equivalent to $N(M+1)$ dimensional coupled Langevin equations:

$$\begin{aligned}
\dot{x}_i \quad &= \frac{1}{V_0}\left[\sum_{s_i} g_{s_i}^i \left(\prod_{s_i^\mu=1}(1-c_i^\mu)\prod_{s_i^\nu=0} c_i^\nu\right) - k_i x_i\right] + \eta_{x_i}, \\
\dot{c}_i^\mu \quad &= \sum_j f_{ij}^\mu (1-c_i^\mu) - \sum_j h_{ij}^\mu c_i^\mu + \eta_{c_i^\mu}
\end{aligned} \tag{3.19}$$

$$\begin{aligned}
\langle \eta_{x_i}(t)\eta_{x_j}(t')\rangle &= \frac{1}{V_0^2}\left[\sum_{s_i} g_{s_i}^i \left(\prod_{s_i^\mu=1}(1-c_i^\mu)\prod_{s_i^\nu=0} c_i^\nu\right) - k_i x_i\right]\delta_{ij}\delta(t-t') \\
\langle \eta_{c_i^\mu}(t)\eta_{c_j^\nu}(t')\rangle &= \left(\sum_j \left(f_{ij}^\mu (1-c_i^\mu) + h_{ij}^\mu c_i^\mu\right)\right)\delta_{\mu\nu}\delta(t-t')
\end{aligned} \tag{3.20}$$

3.2.4 Generalization to gene with multiple transcription level

However, if the genes are regulated also by mechanism such as chromatin structure change [20, 41, 43, 80] or histone modification [81, 82, 83], it should be reasonable to assume that a single gene has more than two states. Then, it is proper to introduce a multiple component spinor:

$$|s_i\rangle = \begin{pmatrix} c_i^1 e^{i\phi_i^1/2} \\ \vdots \\ c_i^M e^{i\phi_i^M/2} \end{pmatrix} \qquad \langle s_i| = \left(e^{-i\phi_i^1/2} \ldots, e^{-i\phi_i^M/2}\right) \tag{3.21}$$

The normalization condition $\sum_\mu c_i^\mu = 1$ is preserved by the form of $\mathcal{H}_i$. In the example of $N = 1$ network where there is one x variable and $(M-1)$ independent c^μ variables. Assume we can still write down the master equation in the coupled FPE form:

$$\partial_t P_s(n,t) = \sum_{s'} (\mathcal{H}_0 + \mathcal{H}_i)_{ss'} P_{s'}(n,t) \tag{3.22}$$

where $\mathcal{H}_0$ is diagonal in s space and $\mathcal{H}_b$ describes the jumping processes. The identity operator in s space would look like:

$$I_s = \prod_{\mu=1}^{M} \left(\frac{1}{4\pi} \int_0^1 dc^\mu \int_{-2\pi}^{2\pi} d\phi^\mu \, |s\rangle \langle s| \right) \tag{3.23}$$

Start with coupled FPE that has Schrodinger Equation like form and follow path integral formula we get the transition probability from initial state $|x_i, s_i, t_i\rangle$ to final state $|x_f, s_f, t_f\rangle$ to be:

$$\begin{aligned} P(x_f, s_f, t_f | x_i, s_i, t_i) = & \quad \langle x_f, s_f | \, e^{\int \mathcal{H} dt} \, |x_i, s_i\rangle \\ = & \quad \text{const} \prod_{i,\mu} \int \mathcal{D}x_i \mathcal{D}p_i \mathcal{D}c_i^\mu \mathcal{D}\phi_i^\mu e^{-\int \mathcal{L} dt} \end{aligned} \tag{3.24}$$

Plug in the identity operator in x space and new one in s space. Similarly we have:

$$\langle p,s,t|1|p,s,t-\Delta t\rangle = 1 + i\sum_i p_i\dot{x}_i\Delta t + i\frac{1}{2}\sum_{i,\mu}\phi_i^\mu\dot{c}_i^\mu\Delta t \tag{3.25}$$

$$\langle p,s,t|\mathcal{H}_0|p,x,t-\Delta t\rangle = \sum_s{}' \left[(\mathcal{H}_0)_{ss}\prod_i c_i(s)\right] \tag{3.26}$$

$$\langle p,s,t|\mathcal{H}_i|p,x,t-\Delta t\rangle = \sum_{s,s'}{}'' \left[(\mathcal{H}_i)_{ss'}\prod_i c_i(s')e^{i(\phi^\nu-\phi^\mu)/2}\right] \tag{3.27}$$

effective Lagrangian turns out to be:

$$\begin{aligned}-\mathcal{L} = i\sum_i p_i\dot{x}_i + \frac{i}{2}\sum_{i,\mu}\phi_i^\mu\dot{c_i^\mu} + \sum_s{}' \left[(\mathcal{H}_0)_{ss}\prod_i c_i(s)\right] \\ + \sum_{s,s'}{}'' \left[(\mathcal{H}_i)_{ss'}\prod_i \left(c_i(s')e^{i(\phi^\nu-\phi^\mu)/2}\right)\right]\end{aligned} \tag{3.28}$$

Here $\sum_s'$ is a sum over all possible s with s_i at the μ state. $\sum_{s,s'}''$ is a sum over all possible s and s' with s_i at the μ state and s_i' at the ν state. As in the $L=2$ case, $\mathcal{H}_0$ alone defines the L^N N-dimensional x space FP landscapes with proper driving force and intrinsic fluctuation. $\mathcal{H}_b$ is non-diagonal . The adiabatic parameter ω can be defined as $\omega = \frac{|\text{typical element of } \mathcal{H}_b|}{|\text{typical element of } \mathcal{H}_0|}$ to reflect the time scale hierarchy of the two sets of processes. We notice that out of the L c^μ elements only $L-1$ are independent due to the probability conservation.

The first order terms, corresponding to deterministic classical dynam-

ics, are:

$$
\begin{aligned}
-\mathcal{L}_{cl} \quad &= i\sum_i p_i\dot{x}_i + \frac{i}{2}\sum_{i,\mu}\phi_i^\mu \dot{c_i^\mu} + \sum_s{}' \sum_i p_i \frac{\partial(\mathcal{H}_0)_{ss}}{\partial p_i}\prod_j c_j(s) \\
+ \quad &\sum_{\substack{s,s' \\ \mu\neq\nu}} \left[(\mathcal{H}_i)_{ss'}\frac{i(\phi^\nu-\phi^\mu)}{2}\prod_i c_i(s')\right]
\end{aligned}
\tag{3.29}
$$

After integration over p_i and ϕ_i^μ, this contributes delta functions which provide deterministic equations:

$$
\begin{aligned}
\dot{x}_i &= i\sum_s{}' \left[\left(\frac{\partial(\mathcal{H}_0)_{ss}}{\partial p_i}\right)\Big|_{p=0}\cdot\prod_j c_j(s)\right] \\
\dot{c_i^\mu} &= \sum_{\substack{s,s' \\ s_i'=\nu\neq\mu=s_i}} \left[(\mathcal{H}_i)_{s's}\prod_j c_j(s) - (\mathcal{H}_i)_{ss'}\prod_j c_j(s')\right]
\end{aligned}
$$

Especially the normalization is conserved at classical level:

$$
\sum_\mu \dot{c_i^\mu} = \sum_{\substack{s,s' \\ s_i\neq s_i'}} \left[(\mathcal{H}_i)_{s's}\prod_i c_i(s) - (\mathcal{H}_i)_{ss}\prod_i c_i(s')\right] = 0 \tag{3.30}
$$

The quadratic terms provide information of intrinsic fluctuation. We are extremely interested in c_i^μ part. Notice that cross terms like $\phi^\mu\phi^\nu$ don't contribute after integration. We are left with quadratic terms of ϕ_i^μ:

$$
\begin{aligned}
\mathcal{L}_{sc} \quad &= \frac{1}{2}\sum_s\sum_i p_i^2\left(\frac{\partial^2(\mathcal{H}_0)_{ss}}{\partial p_i^2}\right) \\
- \quad &\frac{1}{8}\sum_{\substack{s,s' \\ s_i=\mu\neq\nu=s_i'}} (\mathcal{H}_i)_{ss'}(\phi^{\mu 2}+\phi^{\nu 2})\prod_i c_i(s')(\phi^{\mu 2}+\phi^{\nu 2})
\end{aligned}
\tag{3.31}
$$

Just like in $M = 2$ spinor case we have worked out, the quadratic terms, after HubbardStratonovich transformation, provide non-constant gaussian fluctuation:

$$\dot{x}_i = i \sum_s{}' \left[\left(\frac{\partial (\mathcal{H}_0)_{ss}}{\partial p_i} \right) \Big|_{p=0} \cdot \prod_j c_j(s) \right] + \eta_{x_I} \tag{3.32}$$

$$\dot{c_i^\mu} = \sum_{\substack{s,s' \\ s_i'=\nu\neq\mu=s_i}} \left[(\mathcal{H}_j)_{s's} \prod_i c_j(s) - (\mathcal{H}_i)_{ss'} \prod_j c_j(s') \right] + \eta_{c_i^\mu} \tag{3.33}$$

with diffusion coefficients:

$$\langle \eta_{x_i}(t) \eta_{x_j}(t') \rangle = - \sum_s \left[\frac{\partial^2}{\partial_{p_i}^2} (\mathcal{H}_0)_{ss} \left(\prod_j c_j(s) \right) \right] \delta_{ij} \delta(t - t') \tag{3.34}$$

$$\langle \eta_{c_i^\mu}(t) \eta_{c_i^\xi}(t') \rangle = \sum_{\substack{s,s' \\ s_i=\mu\neq\nu=s_i'}} \left[(\mathcal{H}_j)_{s's} \prod_i c_j(s) + (\mathcal{H}_i)_{ss'} \prod_j c_j(s') \right] \delta_{\mu\xi} \delta(t - t') \tag{3.35}$$

Because only $L-1$ of the L c_i^μ variables are independent, those are $N \times L$ dimensional coupled Langevin equations. From the view of probability distribution, this is equivalent to a $N \times L$ dimensional FP landscape system with N-dimensional x space and $N \times (L-1)$ c-space. The evolution in the continuous spinor coordinate c-space plays an important role in the network evolution, especially in the non-adiabatic region. At adiabatic limit, because the dynamics of c variables are totally determined by x-variables, we go back to the N-dimensional x-space FP landscape picture. When $L = 2$, we get back to the on/off N-gene network in the last section.

3.2.5 Gauge Theory

FPE can be written in a Schrodinger Eq. like form. It's very similar to Schrodinger Eq. of a charged particle moving in EM field. We can identify the corresponding scalar field as well as vector field, and specify their gauge transformation rules. The field strength tensor, or curvature in differential geometry, is related to heat dissipation which is a path dependent physical quantity invariant under gauge transformation.[84]

Define differential operator $\nabla_i = \partial_i - \frac{1}{2}D_{ij}^{-1}F_j$, FPE has the Schrodinger Equation like form:

$$\frac{1}{D}\partial_t(D\rho) = \nabla^2(D\rho) - \left[\frac{1}{2}\left(\partial_i\left(\frac{F}{D}\right)_i\right) + \frac{F^2}{4D^2}\right]D\rho \tag{3.36}$$

This is similar to a charged particle moving in electronic-magnetic field with scalar potential:

$$V = -[\frac{1}{2}\partial_i(\frac{F_i}{D}) + \frac{F^2}{4D^2}] \tag{3.37}$$

and vector potential:

$$A_i \;\; = \;\; -(D^{-1})_{ij}F_j \tag{3.38}$$

We have field strength tensor:

$$R_{ij} = [\nabla_i, \nabla_j] = \partial_i A_j - \partial_j A_i \tag{3.39}$$

which is invariant under gauge transformation:

$$A_i \to A_i + \partial_i \Lambda \tag{3.40}$$

Heat dissipation along a closed loop:

$$\begin{aligned} T\Delta s_m &= -\oint_C A_i d_x i = -\oint_C (D^{-1})_{ij} A_j d_x i \\ &= -\frac{1}{2}\int_\Sigma d\sigma_{ij} R_{ij} \end{aligned} \tag{3.41}$$

is a gauge invariant physical quantity.[85]

If $N = 1$, in adiabatic case there's no curvature structure as x-space is a one dimensional space: there're not enough degrees of freedom for vector potential to be non-trivial.

If $N > 1$, in adiabatic limit the landscape and flux live in N dimensional x space. We have a set of N-component vector field, and $\frac{N(N-1)}{2}$ independent field strength tensor components, as defined in Eq. (3.38) and Eq. (3.39).

The field strength tensors are related to heat dissipation as Eq. (3.41) shows. As N-components vector field A_i only depends on x, heat dissipation gets contribution only from evolution in x-space.

But in spin variable formalism, even when N=1, by introducing external dimension c, we have 2-dimensional non-trivial vector potential (A_x, A_c), thus nontrivial curvature:

$$R_{xc} = \partial_x A_c - \partial_c A_x \tag{3.42}$$

$$R_{cx} = \partial_c A_x - \partial_x A_c \tag{3.43}$$

$$R_{xx} = R_{cc} = 0 \tag{3.44}$$

This can be easily generalized to finite N case, with scalar potential defined as Eq. (3.37), vector potential defined as in Eq. (3.38). Because now we are working in 2N dimensional $x - c$ space, vector potential has

2N components (A_x, A_c).

$$A \quad = \quad (A_x, A_c) = (-\frac{F_x}{D_{xx}}, -\frac{F_c}{D_{cc}}) \tag{3.45}$$

The field strength tensor can still be defined as in Eq. (3.39), and is invariant under gauge transformation Eq. (3.40).

Compared with the gauge theory we had before in x-space, the main differences are:

1. Scalar potential contains contribution from c space as well as from x space.

$$V = -[\frac{1}{2}(\partial_x \frac{F_x}{D_{xx}}) + \frac{{F_x}^2}{4{D_{xx}}^2}] - [\frac{1}{2}(\partial_c \frac{F_c}{D_{cc}}) + \frac{{F_c}^2}{4{D_{cc}}^2}] \tag{3.46}$$

2. Vector potential is 2N dimensional.

$$(A_x, A_c) = (-\frac{F_x}{D_{xx}}, -\frac{F_c}{D_{cc}}) \tag{3.47}$$

3. Field strength tensor contains contributions $R_{x_m x_n}$ from x space, $R_{c_m c_n}$ from c space, and cross terms $R_{x_m c_n}$ from $x - c$ space. When calculating heat dissipation of a closed loop, as Eq. (3.39) and Eq. (3.41) shows, even when filed strength tensors in x space and c space vanish, the cross terms can be non-trivial and contributes to the heat dissipation rate.

$$T\Delta s_m = -\oint_C A_i d_x i == -\int_\Sigma d\sigma_{xc} R_{xc} \tag{3.48}$$

4. The phase Δs_m is gauge invariant under gauge transformation Eq. (3.40). The origin of this phase is the non-trivial curvature of internal

space caused by the deviation from equilibrium state. This is similar to Berry phase in quantum phenomenon.

3.2.6 Non-equilibrium thermodynamics and FDT

FPE can be written as:

$$\partial_t \rho = \hat{L}\rho = [-\partial_i F_i + \partial_i \partial_j D_{ij}]\rho \tag{3.49}$$

At steady state, do small perturbation to driving force F:

$$F(x) \to F'(x) = F(x) + h(t)\delta F(x) \tag{3.50}$$

The FPE operator becomes $\hat{L}' = \hat{L} - h(t)\delta\hat{L}$, where $\delta\hat{L} = \partial_i(\delta F_i) + \delta F_i \partial_i$ The probability evolves as:

$$P(x,t) = \exp\left[\int_{t'}^{t} d\tau(\hat{L} - h(\tau)\delta\hat{L})\right] P(x,t') \tag{3.51}$$

For $t \leq t'$, define response function:

$$\begin{aligned} R^{\Omega} &= \left.\frac{<\delta\Omega(t)>}{\delta h(t')(t-t')}\right|_{\delta F=0} && (3.52)\\ &= \left.\frac{\int dx\Omega[\rho(x,t) - \rho_{ss}(x,t)]}{\delta h(t')(t-t')}\right|_{\delta F=0} && (3.53)\\ &= \int dx\Omega(x)e^{\hat{L}(t-t')}(-\delta\hat{L})\rho_{ss}(x) && (3.54) \end{aligned}$$

Plug in the expression we get for δL:

$$\begin{aligned} R^{\Omega} = \quad & - <\Omega(t)\partial_i\delta F_i(t')> && (3.55)\\ & - \left[<\Omega(t)\delta F_i(t')\tilde{F}(t')D^{-1}_{ik}> + <\Omega(t)\delta F_i(t')V^{(ss)}_k(t')D^{-1}_{ik}(t')\right] && (3.56) \end{aligned}$$

If the perturbation is independent of x: $\delta F = 1$, we get:

$$R_i^{\Omega}(t-t') = - < \Omega(t)\partial \ln \rho_{ss}(x) > \quad (3.57)$$

$$= - < \Omega(t)\tilde{F}_k(t')D_{ik}^{-1}(t') > - < \Omega(t)V_k^{ss}(t')D_{ik}^{-1}(t') > \quad (3.58)$$

This is generalized FDT in non-equilibrium thermodynamics.[85] The second term on the right hand side is directly related to the non-equilibrium effects, as we can see it is related to the non-trivial flux $j^{SS} = \rho^{SS}V^{SS}$.

Choose $\Omega = V_i$ and sum over i:

$$R^V = \int dx V(x)[-\partial\rho(x)] = \int dx[\partial \cdot j(x)] \ln \rho(x) \quad (3.59)$$

$$= \frac{d}{dt}\int dx \rho(x) \ln \rho(x) = -\dot{S} \quad (3.60)$$

Entropy changing rate can be decomposed into two parts:

$$\dot{S} =< V_i \cdot \partial_i \ln \rho(x) >=< V_i D_{ij}^{-1} V_j > + < V_i D_{ij}^{-1} \tilde{F}_j >= e_P - \dot{S}_m \quad (3.61)$$

e_P is the entropy production rate of the system, $\dot{S}_m$ is the entropy production rate caused by heat dissipation. These two entropy production rates are closely related to flux V_i.

We want to separate the steady state contribution and non steady state contribution. For this purpose we take $\Omega = V_i - V_i^{SS}$ in the FDT equation:

$$< V_i \partial_i \ln[P^{SS}/P(x)] >= \dot{F}_{free}/T \quad (3.62)$$

$$= < v_i^{SS} {D_{ij}}^{-1} v_j > - < v_i {D_{ij}}^{-1} v_j >= Q_{hk}/T - e_p \quad (3.63)$$

where Q_{hk} is the house keeping heat, $F_{free} = T < \ln[P(x)/P^{SS}(x)] >$ is free energy.

In adiabatic case, the landscape and flux live in x-space. If $N = 1$, x

is one-dimensional, there will be no flux at all. We can write down the entropy production rate:

$$\dot{S} =< V \cdot \partial_x \ln \rho(x) >=< V_i (D_{xx})^{-1}_{ij} V_j > + < V_i (D_{xx})^{-1}_{ij} \tilde{F}_j >= e_P - \dot{S}_m \quad (3.64)$$

We can distinguish entropy production rate from the system and from heat dissipation, of each landscapes.

$$e_P =< V_i (D_{xx})^{-1}_{ij} V_j > \quad (3.65)$$

$$\dot{S}_m = - < V_i (D_{xx})^{-1}_{ij} \tilde{F}_j > \quad (3.66)$$

The house keeping heat and free energy changing rate reads:

$$Q_{hk}/T =< V_i^{SS} (D_{xx})^{-1}_{ij} V_j > \quad (3.67)$$

$$\dot{F}_{free}/T = Q_{hk}/T - e_P \quad (3.68)$$

As in adiabatic case, the flux lives in x space. Thus all these entropy production rates are calculated according to the evolution in x space.

In the spin-variable formalism we workd out which can deal with non-adiabatic case, as flux lives in $2N$ dimensional $x - c$ space, the flux in the non-equilibrium term shall be replaced with new flux in $x - c$ space. There will be new contribution coming from c space, representing jumpings between different states.

The entropy production rate now looks like:

$$\dot{S} =< V_i \cdot \partial_i \ln \rho(x) >=< V_i D^{-1}_{ij} V_j > + < V_i D^{-1}_{ij} \tilde{F}_j >= e_P - \dot{S}_m \quad (3.69)$$

with system and heat dissipation EPRs:

$$e_P =< V_i D_{ij}^{-1} V_j >=< V_x {D_{xx}}^{-1} V_x > + < V_c {D_{cc}}^{-1} V_c > \tag{3.70}$$

$$\dot{S}_m = - < V_i D_{ij}^{-1} \tilde{F}_j >= - < V_x {D_{xx}}^{-1} \tilde{F}_x > - < V_c {D_{cc}}^{-1} \tilde{F}_c > \tag{3.71}$$

The house keeping heat and free energy changing rate now depends on $x - c$ space flux as well.

$$Q_{hk}/T =< V^{SS}_{x(i)} {D_{xx(ij)}}^{-1} V_{x(j)} > + < V^{SS}_{c(i)} {D_{cc(ij)}}^{-1} V_{c(j)} > \tag{3.72}$$

$$\dot{F}_{free}/T = Q_{hk}/T - e_P \tag{3.73}$$

We see under our framework, there are additional contribution terms from c-space to both of the entropy production rates. This is no surprise. With the introduction of the continuous spin variable c, the system evolves in c space as well as in x space. There will be contribution coming from evolution in c space as well as from x space. This is especially meaningful for non-adiabatic case: in non-adiabatic region, jumping rates are comparable to x-space evolution. The adiabatic picture no longer works and we have to deal with jumping in a similar way to x-space evolution. This is what spin variables are introduced for. We see clearly here it brings additional contribution to thermodynamic quantities like entropy production rates.

3.3 Application to self activator

We will apply our theory to a self activating motif with transcription factor being dimer. This is one of the basic building bricks of larger networks.The master equation, in large volume limit, has the form of coupled Fokker-Planck equation as in Eq. (3.2). Introducing two components spinor to quantify the binding site, the equivalent two dimensional Fokker-Planck equation in extended space is:

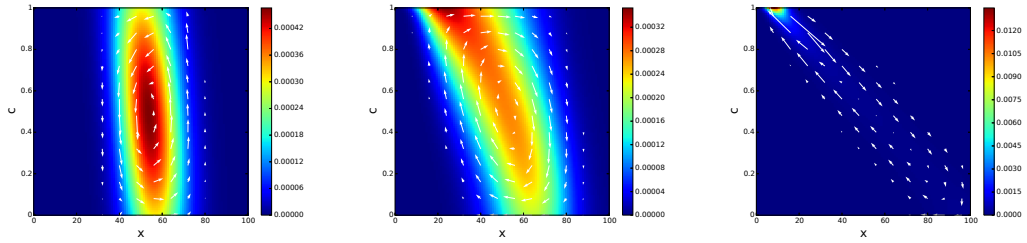
$$\frac{\partial}{\partial t}P = -\partial_x(F_x P) - \partial_c(F_c P) + \partial_x^2(D_x P) + \partial_x^2(D_c P) \tag{3.74}$$

with driving force and diffusion:

$$\begin{aligned} F_x &= \frac{1}{V_0}\left(cg_0 + (1-c)g_1 - kx\right) \\ F_c &= f(1-c) - \frac{1}{2}h_0 x^2 c \\ D_x &= \frac{1}{2V_0^2}[cg_0 + (1-c)g_1 + kx] \\ D_c &= \frac{1}{2}[f(1-c) + \frac{1}{2}h_0 x^2 c] \end{aligned} \tag{3.75}$$

As Fig. 1 suggests, when ω is small, the coupling between on/off state is weak. The landscape shows two separate basins. At adiabatic limit($\omega = 1000$), the two basins merge into one as adiabatic approximation suggests. As Fig. 2 shows, in the new formalism of Fokker-Planck system in extended space, the two separate basins at small ω show up at $c = 0$ 'on' state and $c = 1$ 'off' state. As ω increases, the two basins move against each other and at adiabatic limit coincide at almost the same location, correspond to single peak at adiabatic limit.

The adiabatic approximation that maps the system to a 1-dimensional



(a) $\omega = 1000$ (b) $\omega = 1$ (c) $\omega = 0.01$

Figure 3: Steady state distribution of self-activator in extended space with white arrow being non-gradient flux at (a): $\omega = 1000$, (b):$\omega = 1$, (c):$\omega = 0.01$

Fokker-Planck system, though numerically a good approximation, lacks the ability to demonstrate non-equilibriumness of the system as there is no curl flux in one-dimensional systems. At adiabatic limit, the $\mathbf{K}$ matrix dominates over $\mathbf{L}$. We can construct a normalized steady state solution $\mathbf{r_0}$ of jumping matrix $\mathbf{K}$.

$$\mathbf{H_i r_0} = 0 \tag{3.76}$$

$$\mathbf{r_0} = \frac{1}{\sqrt{h(x)^2 + f(x)^2}} \begin{pmatrix} h(x) \\ f(x) \end{pmatrix} \tag{3.77}$$

We can assume the actual probability $\mathbf{P}$ is $\mathbf{r_0}$ plus a small fluctuation $\mathbf{P} = P\mathbf{r_0} + \xi$.

Equation (8) becomes:

$$\partial_t P = \mathbf{1}^T \mathbf{H_0}(P\mathbf{r_0} + \xi) \approx \mathbf{1}^T \mathbf{H_0}(P\mathbf{r_0}) \tag{3.78}$$

This is a second order PDE of $P(x)$ and has the form of FPE. Which gives a clear landscape picture in x space. In $N = 1$ self regulating case it

looks like:

$$\partial_t P = -\partial_x(\frac{h(x)}{\sqrt{h(x)^2+f(x)^2}}PF_1) + \frac{1}{2}\partial_x^2(\frac{h(x)}{\sqrt{h(x)^2+f(x)^2}}PD_1)$$
$$-\partial_x(\frac{f(x)}{\sqrt{h(x)^2+f(x)^2}}PF_0) + \frac{1}{2}\partial_x^2(\frac{f(x)}{\sqrt{h(x)^2+f(x)^2}}PD_0) \quad (3.79)$$

One-dimensional system is governed by equilibrium gradient driving force and always leads to detailed balance. In our new formalism, as showed in Fig. 3 that the non-trivial flux (white arrows) originated from the spin coupling emerges in the expanded x-c space as a sign of system being at non-equilibrium state.

The Fano Factor of steady states, defined as variance/mean, quantifies the global stability. For a single Poisson peak, Fano Factor equals one. Large fano factor indicates large deviation from single Poisson peak and large fluctuation. Fig. 4(a) shows how Fano factor changes with adiabaticity. In the adiabatic limit, as adiabatic approximation suggests, due to frequent binding/unbinding processes, system is always close to the steady state of $\mathbf{H}_b$. It is a weighted average of 'on' and 'off' peak and is itself Poisson like. The fano factor is close to one. In the moderate non-adiabatic region, Fano factor increase as adiabaticity decreases. It is because when the coupling between 'on' and 'off' states is weak, the adiabatic peak splits into 'on' and 'off' peaks that contribute to Fano factors.

While for 1-dimensional landscape of self activator, the driving force is purely gradient, the transitions from both off to on and on to off are reversible. When extended to 2-dimensional landscape the transition paths are clearly non-reversible with emergence of eddy-current in addition to gradient component. The 2-dimensional landscape makes it possible to study irreversible optimal transitions between off and on states. We focus on the mean first passage time (MFPT) for both on to off and off to on

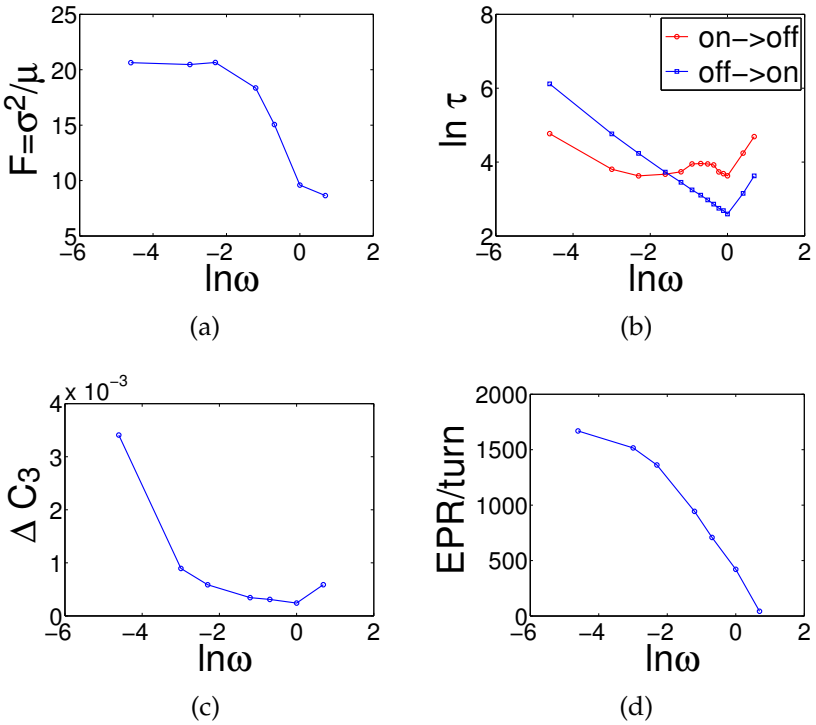


Figure 4: Thermodynamic quantities of self activator calculated in extended space under unified landscape framework (a):Fano factor. (b):Mean first passage time(MFPT). (c):Difference between forward and backward three-point correlation functions. (d):Entropy production rate per turn of the gene on-off switch.

states. From Fig. 4(b) we see a kinetic turnover behavior where an optimal time or transition rate exists with respect to adiabaticity. From mean first passage time, we see that in the adiabatic regime, the discrete landscapes are strongly coupled. The rate limiting step is determined by the adiabatic barrier between the two states on a single effective landscape (intra-landscape dynamics). Decreasing the ω decreases the coupling between discrete landscapes. The 'averaged' single basin at adiabatic limit splits into 'on' and 'off' states as ω decreases. The effective barrier becomes less

than the one in the adiabatic case and the transition rate become higher. On the other hand, in the small ω regime, the rate limiting step is determined by the non-adiabatic jumping between the landscapes. The discrete landscapes are weakly coupled, system tends to stay in either 'on' or 'off' state for a long time. Decreasing the ω further decreases the frequency or probability of jumping and therefore the transition rate decreases. This explains the turnover behavior of rate or kinetic time with respect to adiabaticity ω. [19, 44]

Difference in forward and backward three-point correlation functions, as Fig. 4(c) shows, indicates the system evolution is irreversible in time due to the presence of the flux that breaks detailed balance. As ω decreases (non-adiabaticity increases), the irreversibility becomes more pronounced. We should note that these non-equilibrium features become more evident as ω decreases from the adiabatic regime to the non-adiabatic regime.

The entropy production rate representing the dissipation directly associated with the degree of deviation from equilibrium, is closely related to the new flux in extended space. Fig. 4(d) shows EPR decreases as ω increases. The non-zero flux promotes the non-adiabatic fluctuations and dissipation in the smaller ω regime. Three-point correlation and entropy production rate both show that non-adiabatic regime where the discrete landscapes are weakly coupled shows higher deviation from the equilibrium reflected by the higher heat dissipation and higher irreversibility. Our theory provides a unified landscape theory in this regime with 2-dimensional eddy current that measures the degree of the system being deviated from equilibrium.

Above all, our theory is capable to explain the non-equilibrium and irreversible properties of network dynamics with eddy current emerging in extended space, and provides a theoretical framework to study dynamics (transition rate, path) within a single unified landscape for both adiabatic

and non-adiabatic dynamics.

3.4 Conclusions

In this work, we systematically studied discretely coupled stochastic processes and associated non-adiabatic dynamics that are common in physics and biology. We started with exact master equations, and showed by using the continuous spinor representation we are able to map the dynamics in 2^N discretely coupled landscapes into a single landscape in the extended space. Our result, at deterministic level, agrees with self-consistent approximation. It also unveils the role intrinsic fluctuation played in system evolution. Intrinsic fluctuation lies not only in synthesis and degradation processes, but also in the binding/unbinding processes. In other words, for a coupled Fokker-Planck system (e.g. N-gene network), we have intrinsic fluctuation not only within each of the discrete Fokker-Planck landscapes, but also in the hopping processes. This is clearly demonstrated in the unified landscape formalism with fluctuation in both protein concentration x-variables and discrete state specifying c-variables.

The unified landscape picture provides a theoretical and analytical framework for non-equilibrium dynamics at non-adiabatic region. The mapping from 2^N discretely coupled N-dimensional stochastic system into a $2N$-dimensional Fokker-Planck system significantly improves numerical calculation efficiency in non-adiabatic region where adiabatic approximation is not valid. As we can see from the self activator example, it captures thermal dynamical properties like fluctuation enhancement, irreversibility, transition rate turn over. The steady state flux introduced in the extended space is the key signature of the non-equilibrium dynamics. It gives a quantitative measure of the detailed balance breaking. It also enables us to study optimal transition path and rate under familiar Fokker-

Planck framework.

Chapter 4

Non-adiabatic dynamics as possible source of cancer heterogeneity

4.1 Introduction

Tumor cells are known to have remarkable variability in phenotypes, a phenomenon known as heterogeneity. The distinct phenotypes lead to diversified biological behavior. This is important in cancer research and clinical therapy. The origin of heterogeneity is closely related to the mechanism of cancer [86]. The idea of Darwinian-like clonal evolution is based on the discovery of acquisition of oncogenic mutations. It stresses the gene-centric development in which heterogeneity arises from the diversity of genotypes resulting from clonal evolution. On the other hand, the cancer cells are often hierarchically organized into non tumorigenic and tumorigenic cells with distinct phenotype manifestations [87, 88]. The differences in tumorigenic potential within the same tumor is largely determined by epigenetic diversification. In other words, the origin of hetero-

geneity is often non-genetic[89]. The heterogeneity can be from the epigenetics and micro-environment such as DNA methylation and histone modifications. It can also come from stochastic nature of biological and chemical reactions involved. More and more evidences show that genetic heterogeneity is not likely to make major contribution to cancer heterogeneity [90, 91, 92, 93]. Epigenetic origin of cancer heterogeneity is of both academic and clinical interest.

The cancer was often thought to be determined by the individual gene mutations. More evidences accumulated that cancer is a disease state emerged from the whole gene network, rather than only through individual gene mutations[94]. The results of dynamics and evolution for the gene network are largely determined by the topology of the network while at the same time the gene network dynamics is stochastic with intrinsic fluctuation from statistical molecular number fluctuations and external fluctuation from cell environment[69, 95]. The deterministic and stochastic dynamics of cancer core network can provide possible source for cancer heterogeneity.

In cancer gene regulatory network, a key motif often emerged for function and cell fate decision making is the self activation and mutual repression[14, 96, 97]. For instance Rac1/RhoA circuits mediates amoeboid/mesenchymal transitions in Metastatic carcinoma cells[98]. miR200/ZEB double negative circuits in many cancer cells[99]. To explore heterogeneity arising from network dynamics, we will study one such cancer gene motif.

The key character of the dynamics for this cancer gene motif is the involvement of multiple timescales: timescale of protein synthesis/degradation and timescale of gene regulation or gene state switching processes. Adiabaticity is introduced to quantify the hierarchy of these time scales[14, 19, 20, 38, 39, 43]. Many studies are concentrated in adiabatic limit where the rates for gene state switching due to regulations are much faster than the

rates of protein synthesis/degradation. At this limit, adiabatic approximation is valid and driving force often has the form of Hill function[39, 77]. Two major stable states often emerge in this limit with one gene activated and the other repressed. We identify these two states to be normal state and cancer state. Analytical and numerical studies show that under certain conditions there can be a third intermediate state in addition to normal and cancer states. The intermediate state has the nature of premalignant state [52, 53, 97, 100, 101].

Although adibatic (fast regulation) assumption might be good for some prokaryotic cells such as bacteria, the regulation time scales can be elongated by the epigenetic factors in eukaryotic cells such as DNA methylation and histone remodifications. As M. Sasai and other researchers pointed out, though histone modification of single histones can be quick and frequent, the cooperative change of many histones which represents histone states occurs on a much slower time scale accompanied by dynamical DNA methylation/demethylation [43, 102] When the reaction time scale is extended to non-adiabatic region with slower regulation. The translation/transcription and degradation processes happen more frequently compared to the regulatory processes. The gene states rarely change and there's sufficient time for protein copy number to reach transcription level determined by the gene state. This usually leads to slower epigenetic states change.

In this study we suggest a possible source of cancer heterogeneity as being from epigenetics through gene regulation dynamics in the non-adiabatic (slow regulation) regime. The cancer heterogeneity is reflected from the emergence of more phenotypic states, larger protein concentration fluctuations, wider distribution of kinetics and multiplicity of paths from normal to cancer state, higher energy cost per gene switching and weaker stability.

4.2 Model

We choose our cancer gene regulatory motif with mutually repression and self activation through transcription factor binding/unbinding. Fig.5 shows the regulation scheme of the two genes. Gene A and gene B each has 2 binding sites. The first binding site can be bound to a monomer produced by the other gene and the synthesis rate will be suppressed by a factor λ_R at the bound state (representing the repression gene regulation). The second binding site can be bound to a tetramer produced by itself and the synthesis rate will be raised by a factor λ_A (representing activation gene regulation). In previous similar network models, the form of transcription factor is usually chosen from monomer to tetramer. While there's no direct experiment evidence as of which form transcription factor takes, we make a resonable assumption that TF factors take monomer-tetramer form. The 4 discrete states of each gene has protein synthesis rates set as:$g_{00}, g_{01} = g_{00}\lambda_A, g_{10} = g_{00}\lambda_R, g_{11}\lambda_A\lambda_R$. The first index i in g_{ij} represents the first binding site being at bound (1) state or unbound (0) state, the second index j represents the second binding site. The degradation rate for both proteins is set as $k = 1$. For simplicity, the unbinding rate for all binding sites is set as f. The binding rate for the first binding site (of both gene A and B) is given as $h_{i1} = \frac{f}{Xeq_1}n_j$ and for the second binding site is given as $h_{i2} = \frac{f}{Xeq_2}n_i(n_i-1)(n_i-2)(n_i-3)$. Xeq_1 and Xeq_2 are equilibrium constants(ratio of binding and unbinding kinetics). The adiabatic parameter is defined as $\omega = f/k$.

For the self activating and mutually repressing network motif, the two genes Gene A and Gene B each has two binding sites. The first binding site of gene A(B) can be bound to a monomer produced by gene B(A) and the synthesis rate of protein A(B) will be repressed by a factor λ_R. The second binding site of gene A(B) can be bound to a tetramer produced by

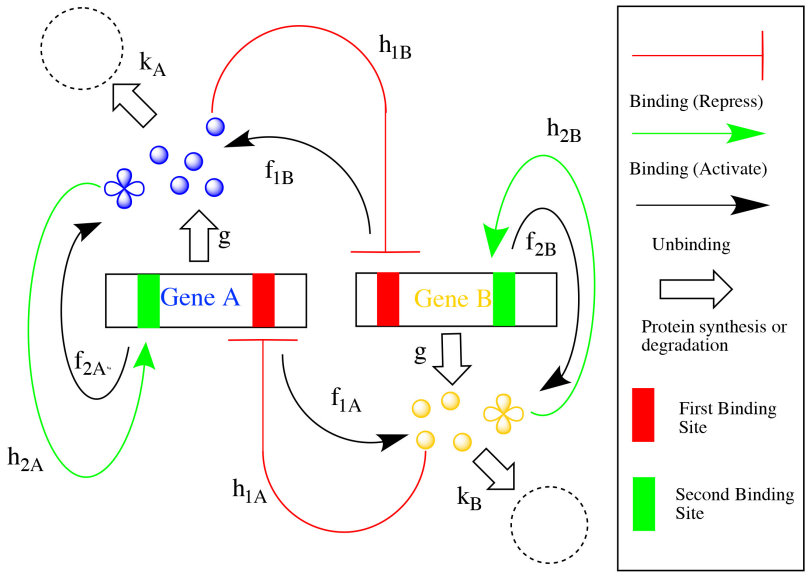


Figure 5: Regulation scheme of self activating mutually repressing regulatory motif

gene A(B) and the synthesis rate of A(B) will be raised by a factor λ_A. We are using multiplicative model with activation and repression effectively measured as multiplicative factors instead of conventional additive factors. We believe multiplicative model is more close in practice as to how gene regulatory network make biological logic decisions. The synthesis rates for different bound/unbound states are $g_{00}, g_{10} = \lambda_R g_{00}$, $g_{01} = \lambda_A g_{00}$, $g_{11} = \lambda_R \lambda_A g_{00}$. The degradation rate for both protein A and protein B are set equat $k_A = k_B = k$. The unbinding rate for all the 4 binding sites are set equat $f_{1A} = f_{2A} = f_{1B} = f_{2B} = f = k \cdot \omega$. The binding rate for the first binding site of gene A(B) is $h_{1A} = \frac{f}{Xeq_1} n_B$ ($h_{1B} = \frac{f}{Xeq_1} n_A$). The binding rate for the second binding site of gene A(B) is $h_{2A} = \frac{f}{Xeq_2} n_A(n_A-1)(n_A-2)(n_A-3)$ ($h_{2B} = \frac{f}{Xeq_2} n_B(n_B-1)(n_B-2)(n_B-3)$). The model can be expressed by

the chemical reactions:

$$\text{Gene}A^{0\alpha} + (n+1)B \underset{h_{1A}}{\overset{f_{1A}}{\leftrightarrows}} \text{Gene}_A^{1\alpha} + nB \qquad \text{Gene}A^{\alpha 0} + (n+4)A \underset{h_{2A}}{\overset{f_{2A}}{\leftrightarrows}} \text{Gene}A^{\alpha 1} + nA$$

$$\text{Gene}B^{0\alpha} + (n+1)A \underset{h_{1B}}{\overset{f_{1B}}{\leftrightarrows}} \text{Gene}_B^{1\alpha} + nA \qquad \text{Gene}B^{\alpha 0} + (n+4)B \underset{h_{2B}}{\overset{f_{2B}}{\leftrightarrows}} \text{Gene}B^{\alpha 1} + nB$$

$$nA \underset{g}{\overset{k}{\leftrightarrows}} (n+1)A \qquad nB \underset{g}{\overset{k}{\leftrightarrows}} (n+1)B$$

In above reactions, α represents an arbitrary state of the binding site, it can be 0 or 1. At large volume limit, protein concentration $x = \frac{n}{V}$ becomes continuous variable. For simplicity we absorb the volume 'V' into g, f and Xeqs. We set synthesis rates $g_{00} = 5$, $\lambda_A = 8$, $\lambda_R = 0.2$, degradation rate $k = 0.1$, unbinding rate $f = k \cdot \omega$ and binding rate $h_{1A} = \frac{f}{Xeq_1} x_B$ ($h_{1B} = \frac{f}{Xeq_1} x_A$), $h_{2A} = \frac{f}{Xeq_2} x_A^4$ ($h_{2B} = \frac{f}{Xeq_2} x_B^4$). Equilibrium constant for mutual repression is set as $Xeq_1 = 15$ and for self activation $Xeq_2 = 50^4$. The master equation that govers network dynamics has the form of coupled Focker-Planck equation:

$$\partial_t \mathbf{P} = (\mathbf{H_0} + \mathbf{H_b})\mathbf{P} \tag{4.1}$$

$\mathbf{P}$ is a 16 component vector whose component $P_s(x,t)$ represents the probability of the system being at gene state s with protein concentration x at time t.

$$\mathbf{P} = \begin{pmatrix} P_{1111}(x,t) \\ P_{1110}(x,t) \\ P_{1101}(x,t) \\ P_{1100}(x,t) \\ \vdots \\ P_{0000}(x,t) \end{pmatrix} \tag{4.2}$$

$\mathcal{H}_0$ describes protein synthesis and degradation processes. It is diagonal with 16 diagonal elements. Each describes a continuous landscape at corresponding discrete gene state.

$$\mathcal{H}_0 = \begin{pmatrix} \mathcal{L}_{1111} & 0 & 0 & 0 & \dots & \vdots \\ 0 & \mathcal{L}_{1110} & 0 & 0 & \dots & \vdots \\ 0 & 0 & \mathcal{L}_{1101} & 0 & \dots & \vdots \\ 0 & 0 & 0 & \mathcal{L}_{1100} & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & \dots & \mathcal{L}_{0000} \end{pmatrix} \tag{4.3}$$

Each diagonal element $\mathcal{L}_{ijkl}$ (each index i, j, k, l can be either 0 or 1) is a 'Focker-Planck' operator:

$$\mathcal{L}_{ijkl} = -\partial_{x_A}(g_{ij} - k_A x_A) - \partial_{x_B}(g_{kl} - k_B x_B) + \frac{1}{2}\partial^2_{x_A}(g_{ij} + k_A x_A) + \frac{1}{2}\partial^2_{x_B}(g_{kl} + k_B x_B) \tag{4.4}$$

On the other hand, $\mathcal{H}_b$ describes binding/unbinding processes and is non-diagonal. It describes the 'coupling' between discrete gene states.

$$\mathcal{H}_b = \begin{pmatrix} r_{1111} & h_{2B} & h_{1B} & 0 & \dots & \vdots \\ f_{2B} & r_{1110} & 0 & h_{1B} & \dots & \vdots \\ f_{1B} & 0 & r_{1101} & h_{2B} & \dots & \vdots \\ 0 & f_{1B} & f_{2B} & r_{1100} & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & \dots & r_{0000} \end{pmatrix} \tag{4.5}$$

The diagonal elements r_{ijkl} in $\mathcal{H}_B$ are escaping rates from gene state

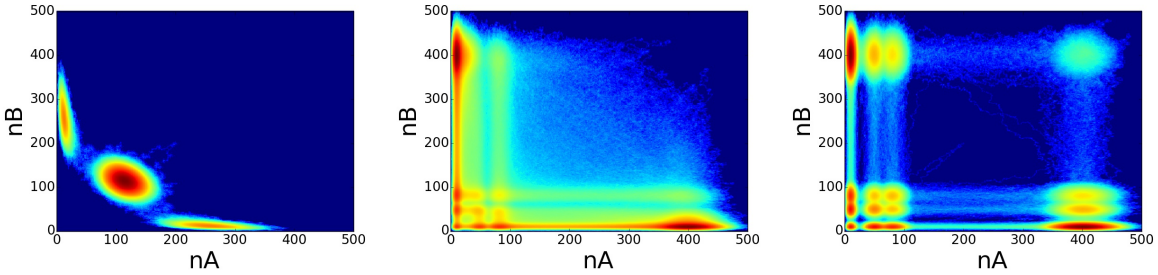
$ijkl$.

$$\begin{aligned} r_{1111} &= -f_{1A} - f_{2A} - f_{1B} - f_{2B} \\ r_{1110} &= -f_{1A} - f_{2A} - f_{1B} - h_{2B} \\ r_{1101} &= -f_{1A} - f_{2A} - h_{1B} - f_{2B} \\ r_{1100} &= -f_{1A} - f_{2A} - h_{1B} - h_{2B} \\ \vdots &= \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ r_{0000} &= -h_{1A} - h_{2A} - h_{1B} - h_{2B} \end{aligned} \tag{4.6}$$

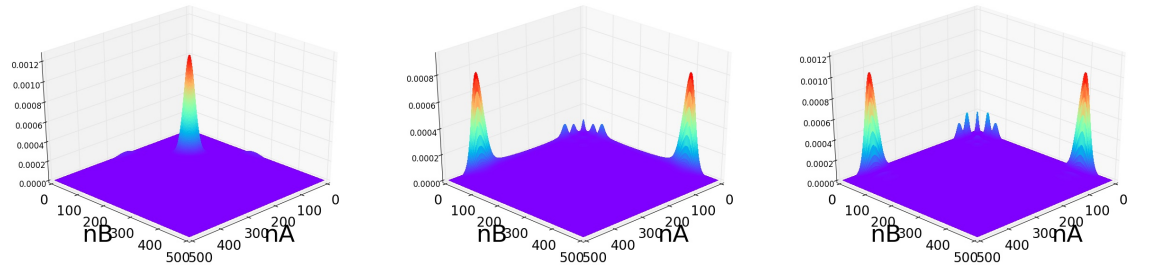
Conventional studies for gene regulation dynamics are concentrated at adiabatic limit of fast regulation binding/unbinding compared to the protein synthesis/degradation. In this case, adiabatic approximation is valid and underlying stochastic dynamics can be described by the dynamic master equation, which has the form of 2-dimensional Fokker-Planck equation in the continuous limit[39]. Deterministic part of the driving force has the form of Hill function[97] while the stochastic part of the force can come from the intrinsic statistical number fluctuations in concentrations or external fluctuations. The problem is greatly simplified. In the more general case, we should use master equation explicitly to describe the stochastic dynamics. Since each gene has four discrete states (00, 01, 10 and 11). The master equation which governs the dynamics, is $2^4 = 16$ dimensional. At large volume limit, the protein concentration variables $x_i = n_i/V$ become continuous. The master equation becomes 'Coupled Fokker-Planck' equation. It has 16 discrete 'Fokker-Planck' states that correspond to the system being at one of the 16 gene states. These Fokker-Planck states are coupled by the binding/unbinding reactions. There are two crucial time scales involved in such system: the timescale of protein synthesis/degradation and the timescale of binding/unbinding of regulatory gene network. Adiabatic parameter $\omega = f/k$ as the ratio between protein regulation bind-

ing/unbinding rate to the gene and the protein synthesis/degradation rate is introduced to quantify the hierarchy of the two timescales. The probability evolution follows coupled Fokker-Planck equation. In the coupled Fokker-Planck equation (master equation in large volume limit), the population or probability $\mathbf{P}$ is a 16 component state vector. Each component P_s stands for the probability of the system with protein concentration $\mathbf{x}$ being at gene state 's'. $\mathbf{H_0}$ and $\mathbf{H_b}$ are operators that can act on $\mathbf{P}$. $\mathbf{H_0}$ is diagonal with 16 operators that describe the protein synthesis and degradation processes. $\mathbf{H}_b$ is non-diagonal with binding/unbinding terms that describe the 'coupling' between the gene states. The physical picture of coupled system is clear. Each component of $\mathbf{H_0}$ defines a probability landscape corresponding to a discrete state of the four binding sites being at a specific bound/unbound gene state. $\mathbf{H_b}$ describes the 'hopping' processes between these states. The adiabatic parameter ω is a measure of the relative strength of regulatory binding/unbinding of protein to the gene and protein synthesis and degrdation. When ω is large, $\mathbf{H_b}$ dominates over $\mathbf{H_0}$, the 'hopping' happens so frequently that the 'hopping' processes reaches an equilibrium and adiabatic regime of fast regulation is reached. When ω is small, the 'hopping' is rare and the system tends to stay in one of the 16 gene state landscapes. The non-adiabatic regime of slow regulation is reached. The moderate ω non-adiabatic regime is of special interest where proper approximation or analytical treatment is often lacking.

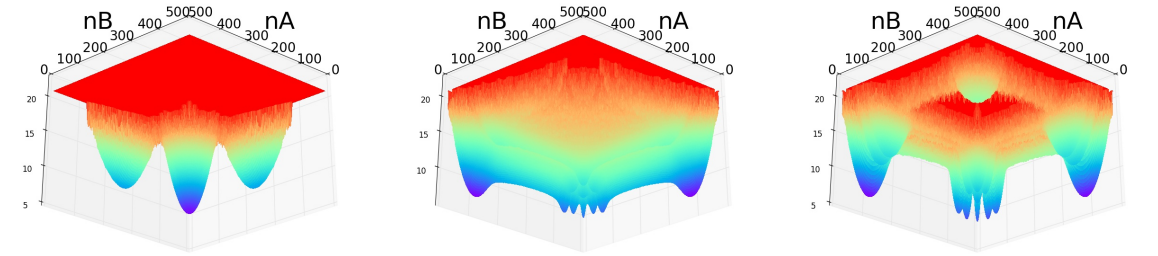
In practice, we use Gillespie algorithm[103] to solve the coupled master equation/Fokker-Planck equations we obtain the steady state probability which gives the underlying landscape of the network system quantifying the probability of the emergent phenotypic states and the evolution dynamics in time covering all the regimes of interests from adiabatic to non-adiabatic case and in between. Various analytical approximations made in adiabatic and non-adiabatic regimes are consistent with the full Gille-



(a) Two dimensional steady state distribution at $\omega = 100$ (b) Two dimensional steady state distribution at $\omega = 0.1$ (c) Two dimensional steady state distribution at $\omega = 0.001$



(d) Three dimensional steady state distribution at $\omega = 100$ (e) Three dimensional steady state distribution at $\omega = 0.1$ (f) Three dimensional steady state distribution at $\omega = 0.001$



(g) Three dimensional landscape at $\omega = 100$ (h) Three dimensional landscape at $\omega = 0.1$ (i) Three dimensional landscape at $\omega = 0.001$

Figure 6: Two and Three dimensional landscape from fast regulation adiabaticity to slow regulation non-adiabaticity at $\omega = 100, 0.1, 0.001$

spie simulations. We also developed a method for quantifying the optimal path in both adiabatic and non-adiabatic regime. Details are included in the supporting information.

4.3 Adiabatic and Non-Adiabatic Dynamics

The gene network dynamics can be simulated with the Gillespie algorithm[103]. Various analytical approaches can be developed to approximate the adiabatic fast regulation and non-adiabatic slow regulation limit. When ω is large, we are at adiabatic limit where adiabatic approximation of fast regulation of protein to the gene compared to the protein synthesis/degradation is valid[39, 77].

When $\mathcal{H}_b$ terms representing binding/unbinding or gene switches dominate over H_0 terms representing protein synthesis/degradation, the dynamics of the fast degree of freedom (gene switching) quickly reaches equilibrium which is the steady state of $\mathcal{H}_b$:

$$\mathbf{H}_b \xi = 0 \tag{4.7}$$

here ξ is a 16 components state vector and is normalized as $\mathbf{1}^T \cdot \xi = \sum_i \xi_i = 1$. Being close to ξ, the total probability distribution can be written as:

$$\mathbf{P} = \rho\xi + \epsilon \tag{4.8}$$

where ρ is a scaler function of x and t, ϵ is a small deviation from $\rho\xi$.

The probability of the gene network with x protein concentration is the sum over the probabilities of the network at x protein concentration but being at different gene states . That is $\mathbf{1}^T \cdot \mathbf{P} = \sum_i P_i \approx \rho$. Multiply the master equation using $\mathbf{1}^T$ from the left, and notice we always have $\mathbf{1}^T \mathbf{H}_b = 0$. The master equation becomes:

$$\partial_t \rho = \mathbf{1}^T H_0 \rho \xi = \sum_i (H_0)_{ii} (\xi_i \rho) \tag{4.9}$$

The right hand side is the sum of 16 Fokker-Planck equations resulting

the final form of 2-dimensional Fokker-Planck equation. The driving force has the form of Hill functions and intrinsic fluctuation is x dependent.

$$\frac{d}{dt}\rho = -\partial_{x_A}(F_A\rho) - \partial_{x_B}(F_B\rho) + \frac{1}{2}\partial^2_{x_A}(D_A\rho) + \frac{1}{2}\partial^2_{x_B}(D_B\rho) \tag{4.10}$$

with driving force and diffusion coefficients:

$$\begin{aligned}
F_A = \quad & g_{00}\frac{f}{f+h_{1A}}\frac{f}{f+h_{2A}} + g_{01}\frac{f}{f+h_{1A}}\frac{h_{2A}}{h_{2A}+f} \\
& + g_{10}\frac{h_{1A}}{h_{1A}+f}\frac{f}{h_{2A}+f} + g_{11}\frac{h_{1A}}{h_{1A}+f}\frac{h_{2A}}{h_{2A}+f} - kx_A \\
F_B = \quad & g_{00}\frac{f}{f+h_{1B}}\frac{f}{f+h_{2B}} + g_{01}\frac{f}{f+h_{1B}}\frac{h_{2B}}{h_{2B}+f} \\
& + g_{10}\frac{h_{1B}}{h_{1B}+f}\frac{f}{h_{2B}+f} + g_{11}\frac{h_{1B}}{h_{1B}+f}\frac{h_{2B}}{h_{2B}+f} - kx_B
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
D_A = \quad & g_{00}\frac{f}{f+h_{1A}}\frac{f}{f+h_{2A}} + g_{01}\frac{f}{f+h_{1A}}\frac{h_{2A}}{h_{2A}+f} \\
& + g_{10}\frac{h_{1A}}{h_{1A}+f}\frac{f}{h_{2A}+f} + g_{11}\frac{h_{1A}}{h_{1A}+f}\frac{h_{2A}}{h_{2A}+f} + kx_A
\end{aligned}$$

$$\begin{aligned}
D_B = \quad & g_{00}\frac{f}{f+h_{1B}}\frac{f}{f+h_{2B}} + g_{01}\frac{f}{f+h_{1B}}\frac{h_{2B}}{h_{2B}+f} && (4.12) \\
& + g_{10}\frac{h_{1B}}{h_{1B}+f}\frac{f}{h_{2B}+f} + g_{11}\frac{h_{1B}}{h_{1B}+f}\frac{h_{2B}}{h_{2B}+f} + kx_B && (4.13)
\end{aligned}$$

When the whole system reaches steady state (protein concentrations in addition to the gene switchings), we can quantify the landscape as :

$$U = -\ln \rho_{SS}(x) \tag{4.14}$$

The stable states correspond to local minimum of the landscape U. The stability of the states is related to the topography of the landscape in terms of basin depths. The Fokker-Planck equation can also be written in

terms of probability conservation with lefthand side being the probability change in time and righthand side as the divergence of the flux. The physical meaning is clear. The probability increase and decrease is associated with the flux in and out. The flux is given as $j_i^{SS} = -F_i\rho^{ss} + \frac{1}{2}\partial_j(D_{ij}\rho^{ss})$ and is divergent free $\nabla \cdot \mathbf{j} = 0$ in steady state. Define $\hat{F}_i = F_i - \frac{1}{2}(\partial_j D_{ij})$, flux can be written as $\mathbf{j} = -\mathbf{F}\rho^{SS} + \frac{1}{2}\partial \cdot (\mathbf{D}\rho^{SS}) = -\hat{\mathbf{F}}\rho^{SS} + \frac{1}{2}\mathbf{D} \cdot \partial\rho^{SS}$, which is similar to constant diffusion case:. The driving force can be decomposed into a gradient part plus a curl part: $F_i = -\frac{1}{2}D_{ij}\partial_j U + j_i^{ss}/\rho^{ss} + \frac{1}{2}\partial_j D_{ij}$ or $\hat{F}_i = -\frac{1}{2}D_{ij}\partial_j U + j_i^{ss}/\rho^{ss}$. Non-zero flux is a measure of how far the system deviates from the detailed balance (how far away the system is from equilibrium). The non-zero flux is the origin of the irreversibility and entropy production[23, 52].

As shown in Fig. 6, simulations under adiabatic approximation confirms that besides normal state and cancer state, an intermediate state emerges as possible source of phenotype alternation. With the dynamic equation being 2-dimensional Fokker-Planck equation under adiabatic approximation, it is possible to quantify optimal path using minimal action approach. The 2D Fokker-Planck system is equivalent to a Lagrangian dynamic system with effective Lagrangian[24, 67]:

$$\mathcal{L} = -\sum_{i,j} \frac{1}{2D_{ij}}(\dot{x}_i - F_i)(\dot{x}_j - F_j) + \sum_{i,j,k} \frac{1}{2} D_{ik}\partial_k(F_j D_{ij}^{-1}) \tag{4.15}$$

Path integral tells us that, among all the possible transition paths connecting initial (normal) state and final (cancer) state, the possibility of a single path is proportional to the exponential of the negative action $S = \int_{t_i}^{t_f} \mathcal{L}dt$. The optimal path is the one that minimizes action S[78].As have been pointed out, by using kinetic equation and energy conservation along optimal path, the action can be simplified into a line integral along the path[104, 105]

$$S = \int_{x_i}^{x_f} dl \sqrt{2(E - V_{eff}(x))} - \int_{l_i}^{l_f} dl \sum_{i,j} F_l \tag{4.16}$$

D_{ij}^{-1} plays the role of metric in the curved space. $dl = \sqrt{\sum_{i,j} D_{ij}^{-1} dx_i dx_j}$ is the 'distance'. F_l is the force projected to the path line modularized by the diffusion. The second term in the action suggests the path is irreversible since the force is not a pure gradient. The curl flux component of the force will lead to non-trivial contribution to the dynamics. This confirms the irreversibility of cancer model: the reversed optimal path will no longer be the optimal path as the new action is larger and the probability for the reversed path is smaller.

Fig. 9 shows clearly that at adiabatic limit, the optimal transition path going through the intermediate state. In other words, a large portion of the transition paths will go through intermediate state, resulting in a hub with portion of the cells showing the intermediate alternative phenotype.

As ω decreases, binding/unbinding processes are less frequent and we reach the non-adiabatic regime. Adiabatic approximation is no longer valid. We have to use 16 dimensional master equation explicitly.

In the small ω, or weak coupling limit. $\mathcal{H}_b$ coupling terms are much smaller than $\mathcal{H}_0$ terms. System tends to stay in one of the 16 states and rarely jumps to another. In this limit, besides normal state and cancer state, we are able to identify more intermediate states. Especially the 'off-off' state where both genes are repressed. These states have a great impact on the cancer transition behaviour. There are also smaller peaks emerging around the cancer state, suggesting possible phenotype alternations even after the system reaches cancer state.

The moderate ω non-adiabatic region is of special interest. Conventionally in this regima solid theoretical framework and analytical results are hard to achieve. Recent progress[24, 66, 72, 79] showed that the sim-

ilarity between coupled Fokker-Planck equation and Schrodinger equation. One can represent Fokker-Planck equation in the operator form and introduce spinor representation to quantify discrete gene states. In other words, the coupled discrete (gene state) and continuous (protein degrees of freedom) processes are very hard to study analytically. However, we can map the problem to the total continuous representation so that analytical approach is possible. This is realized by monitoring the occupation or probability of the gene on and off states instead of the discrete labeling itself, since occupation is a continuous variable. The result is then the continuous processes in extended dimensions (new continuous variables including the occupations of the gene on and off states). Path integral can further map the Hamiltonian dynamics into Lagrangian dynamics with effective Lagrangian. After Hubbard-Stratonovich transformation the effective Lagrangian provides information of the dynamics at both classic (deterministic level) and semi-classical (intrinsic fluctuation) level in the extended continuous space.

In our case, there are 4 on/off binary binding sites, it is proper to introduce 4 two-components spinors. 4 auxiliary variables: $(c_{A0}, c_{A1}, c_{B0}, c_{B1})$ and their conjugate variables are introduced to quantify the dynamics of the spinors. Follow the path integral procedure with spinor representation, the original coupled Focker-Planck system is mapped to a single Fokker-Planck equation in $2 + 4 = 6$ dimensional extended space.

$$\begin{aligned}\partial_t P = \quad & -\sum_{i=1}^{i=2} \partial_{x_i}(\tilde{F}_{x_i} P) - \sum_{j=1}^{j=4} \partial_{c_j}(\tilde{F}_{c_j} P) \\ & + \sum_{i=1}^{i=2} \frac{1}{2} \partial_{x_i}^2 (\tilde{D}_{x_i} P) + \sum_{j=1}^{j=4} \frac{1}{2} \partial_{c_j}^2 (\tilde{D}_{c_j} P)\end{aligned}$$

with driving force and diffusion coefficients:

$$
\begin{aligned}
\tilde{F}_{x_A} &= c_{A0}c_{A1}g_{00} + c_{A0}(1-c_{A1})g_{01} \\
&\quad +(1-c_{A0})c_{A1}g_{10} + (1-c_{A0})(1-c_{A1})g_{11} - k\cdot x_A \\
\tilde{F}_{x_B} &= c_{B0}c_{B1}g_{00} + c_{B0}(1-c_{B1})g_{01} \\
&\quad +(1-c_{B0})c_{B1}g_{10} + (1-c_{B0})(1-c_{B1})g_{11} - k\cdot x_B \\
\tilde{F}_{c_{A0}} &= f(1-c_{A0}) - \frac{f}{Xeq_1}x_B c_{A0} \\
\tilde{F}_{c_{A1}} &= f(1-c_{A1}) - \frac{f}{Xeq_2}x_A c_{A1} \\
\tilde{F}_{c_{B0}} &= f(1-c_{B0}) - \frac{f}{Xeq_1}x_A c_{B0} \\
\tilde{F}_{c_{B1}} &= f(1-c_{B1}) - \frac{f}{Xeq_1}x_B c_{B1}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{D}_{x_A} &= c_{A0}c_{A1}g_{00} + c_{A0}(1-c_{A1})g_{01} \\
&\quad +(1-c_{A0})c_{A1}g_{10} + (1-c_{A0})(1-c_{A1})g_{11} + k\cdot x_A \\
\tilde{D}_{x_B} &= c_{B0}c_{B1}g_{00} + c_{B0}(1-c_{B1})g_{01} \\
&\quad +(1-c_{B0})c_{B1}g_{10} + (1-c_{B0})(1-c_{B1})g_{11} + k\cdot x_B \\
\tilde{D}_{c_{A0}} &= f(1-c_{A0}) + \frac{f}{Xeq_1}x_B c_{A0} \\
\tilde{D}_{c_{A1}} &= f(1-c_{A1}) + \frac{f}{Xeq_2}x_A c_{A1} \\
\tilde{D}_{c_{B0}} &= f(1-c_{B0}) + \frac{f}{Xeq_1}x_A c_{B0} \\
\tilde{D}_{c_{B1}} &= f(1-c_{B1}) + \frac{f}{Xeq_1}x_B c_{B1}
\end{aligned}
$$

The steady state probability distribution provides a global landscape picture. Compared with coupled 16 discrete landscapes, a global unified

landscape with continuous variables in the extended dimensions captures the non-equilibrium property of the system. Non zero flux emerges in extended space as a measure of the system being at non-equilibrium and irreversible. It also provides a clear picture not only for a unified landscape but also for dynamics covering both intra continuous and inter discrite landscapes dynamics. We can see from this formalism that deterministic force and intrinsic fluctuation lie not only within each discrete landscape in x-space, but also in the hopping processes that are quantified by continuous c-variables. The dynamics in extended space can influence the attractors on the landscape and fluctuation induced transitions.

In particular, the optimal transition path can be quantified under this unified framework in continuous extended space. Fig. 10 shows the optimal path obtained using minimal action approach under 6-dimensional unified landscape projected to 2-dimensional $n_A - n_B$ space. Compared with optimal path at adiabatic limit, it's more tilted towards the 'off-off' state. This is due to the location change of intermediate state as a result of the weakened coupling, the change of gradient force of the landscape and non-equilibrium flux.

4.4 Results and Discussions

4.4.1 Cancer Heterogeneity from Landscape View

In our model, there are two time scales quantifying the interaction strengths of the gene networks. One is the protein synthesis (g) and degradation rates (k), while the other is the binding(h)/unbinding(f) of regulatory proteins to the gene leading the gene state to switch. When the relative time scale $\omega = f/k$ is small, then the regulation processes are relatively slow and the gene switch (on and off) slower than the protein synthesis and

degradation. In this non-adiabatic regime, the couplings among genes through protein regulations are loose due to the weak regulations. The individual genes can be either switched on or off without much influences from others. In this non-adibatic regime, the number of states one expects from the gene network can reach up to 2^N where N is the number of genes. When the relative time scale $\omega = f/k$ is large, then the regulation processes are fast and the genes switch faster than the protein synthesis and degradation. In this adiabatic regime, the couplings among genes are tight due to the strong regulations. The on and off states of genes are controlled by the interactions with other genes. Therefore, one expects that only finite number of states emerge as a result of gene interactions.

Due to the intrinsic and extrinsic fluctuations, the gene network dynamics is stochastic. Following the individual stochastic trajectory will not provide global information. We explore the probability evolution instead. The steady state probability landscape quantifies the chances of each individual state. It gives a global description for the network system. From the simulations of the kinetic processes involved in our cancer gene network motif, we see the steady state probability (P_{ss}) and potential landscape ($U = -lnP_{ss}$) projected in two and three dimensions with respect to two protein concentrations (gene products). Heterogeneity is directly related to the number of attractors of the landscape. The numbers and locations of the attractors determine the possible phenotypes that can be observed[23, 52].

As we can see from Fig.6 in main text, the landscape at adiabatic limit, extreme non-adiabatic limit and moderate non-adiabatic regime are very different. In the extreme non-adiabatic region, $\mathbf{H}_b$ terms are much smaller than $\mathbf{H}_0$ terms. The gene states defined by $\mathbf{H_0}$ are decoupled and we can expect as many discrete states as the dimension of $\mathbf{H_0}$ matrix. In our examples, it is up to 16 states. In the other limit where $\mathbf{H}_b$ dominates

over $\mathbf{H_0}$, adiabatic approximation is valid and the system is mapped to an N-dimensional network whose number of stable states tends to be less. The moderate non-adiabatic regime, as we can see, is different from either limit. The location of the stable states are close to the ones in the extreme non-adiabatic regime, but the smaller basins that are close to larger basins merge to major peaks which is similar to the case in adiabatic limit. In the adiabatic regime when the gene network is strongly coupled (gene regulation time scale is much faster than the protein synthesis and degradation), three states emerge. Two states are mutually repressive to each other. One state has one gene on and the other gene off. This leads to the state with one gene expression high and the other gene expression low. Here gene expression is represented by the corresponding protein concentration. The other state is just the opposite. The high expression of the first gene and low expression of the second gene can be used to represent the cancer state, while the high expression of the second gene and low expression of the first gene can be used to represent the normal state. Then we see both normal and cancer states emerge from the gene network motif. Furthermore, both states are quantified by the two basins of attractions on the landscape with large probability. The reason of the normal and cancer state/basin appearance lies in the fact of the mutual repressing interactions among the genes. We also notice that a third state quantified by the basin of attraction emerges representing an gene on-gene on state. The appearance of this state/basin is from the self activation of both genes. Both experimentally and clinically, premalignant state between normal and cancer have been observed[106, 107]. The emergence of intermediate state/basin between normal and cancer state in this core cancer gene regulatory motif may shed lights on quantifying the premalignant state.

As shown in Fig. 6(b), 6(c), 6(e), 6(f), 6(h), 6(i), in the non-adiabatic regime when the gene network is weakly coupled (gene regulation time

scale is slower than the protein synthesis and degradation), more states quantified by local basins of attractions emerge. As mentioned, the slow regulations can come from epigenetic effects of extra time scales from DNA methylation and histone regulations. The normal and cancer state basins still persist under slow regulation regime. Due to the weak binding/unbinding, the self activation is suppressed. The intermediate basin is shifted towards the gene-off gene-off state and splits into multiple states. Besides these three major state basins (normal, cancer and off-off state), we have observed the emergence of many local state basins around the major basins. The multiple local basins around the major basins give a quantitative picture of the cancer heterogeneity. Cancer state is not an individual state but composed of many states as we can see clearly under the epigenetic conditions. The possible physical mechanism of the cancer heterogeneity is thus from the weakening of the regulatory interactions among genes in the gene network. The epigentics such as DNA mythylation and histone remodification can naturally lead to the effective weakening of the gene interactions.

From another angle, we can see as the gene network is weakly coupled due to the weakening of interactions in $\mathbf{H}_b$ of Eq. (4.1). (with epigenetics of DNA methylation and histone remodification being possible source), one expects to have the emergence of many states or phenotypes. The state basins are relatively shallow. Thus this is a possible physical mechanism of cancer heterogeneity. As the coupling among genes become stronger, more and more states are merged together. Shallower basins are merged to deeper and larger basins. As a result, there are less phenotypic states but relatively more stable. While the weak coupling between genes (or landscapes of fixed gene states) at non-adiabatic regime can explain the emergence of alternative phenotypic states, we will do a detailed study of fluctuation, stableness, transitions and thermodynamic properties to fur-

ther unveil the relationship between heterogeneity and non-adiabatic network dynamics.

4.4.2 Cancer Heterogeneity from Fluctuations in Concentrations and Kinetics

We can further characterize the cancer heterogeneity from the fluctuations in concentrations and kinetics. Fig. 7 shows the Fano factor which is the ratio of variance versus the mean of the protein concentrations (gene products). In a pure random process, statistical fluctuations from intrinsic noise has a Poisson nature in which case Fano factor is equal to 1. We see when the adiabaticity parameter ω of relative regulatory binding/unbinding to progein synthesis and degradation is large, the fluctuations is non-Poisson in this adiabatic regime. This is due to the nature of underlying reactions of protein synthesis and degradation. On the other hand, when the adiabaticity parameter ω of relative regulatory binding/unbinding to protein synthesis and degradation is small, the fluctuations are significantly larger. The significant larger fluctuations give the variances and heterogeneity. This heterogeneity is coming from the emergence of the many local state basins of attractions.

In Fig. 8(a), 8(b), 8(c), we also show the heterogeneity in kinetics. We plot the statistical distribution of first passage time from normal to cancer state. We see when the adiabaticity parameter ω of relative binding/unbinding to synthesis and degradation is large (Fig. 8(a)), the distribution of kinetics is rather narrow in this adiabatic regime. This is due to the limited kinetic paths between normal and cancer state basins. On the other hand, when the adiabaticity parameter ω of relative binding/unbinding to synthesis and degradation becomes smaller (Fig. 8(b)), the statistical distribution of the kinetics is significantly wider. The significant wider distribution indicates that there are many more pathways from

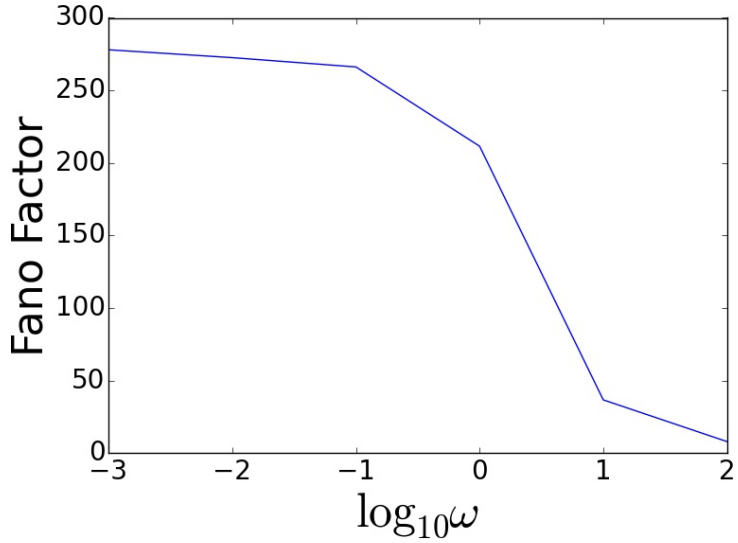
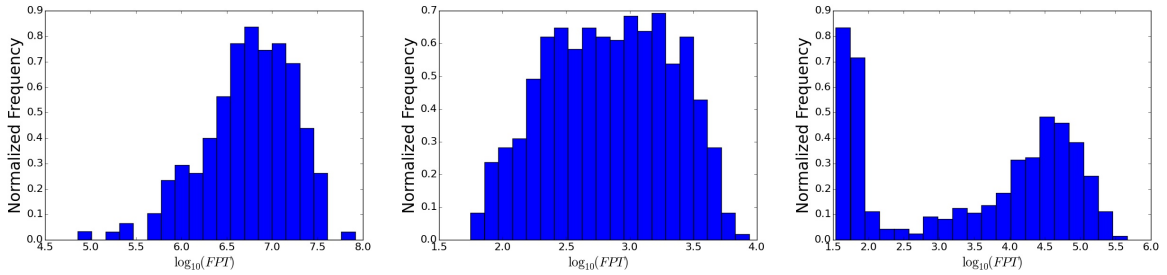


Figure 7: Fano factor from slow regulation non-adiabaticity to fast regulation adiabaticity

normal to cancer state basins. This shows the kinetic variance and heterogeneity. Again, this kinetic heterogeneity is coming from the emergence of the many local state basins of attractions. These new basins of attractions lead to a rougher landscape and therefore multiple non-equivalent pathways from normal to cancer state. Finally, when the adiabaticity parameter ω of relative binding/unbinding to synthesis and degradation is extremely small (Fig. 8(c)), the statistical distribution of the kinetics becomes narrower. Under this condition, the rate limiting step is the switching speed of the genes rather than the protein synthesis or degradations. Due to the slowness of the switching speed caused by the slow regulation, single or limited switchings dominate the kinetics. Therefore, although there are more local basins, there are only limited explorations under extreme slow switchings and the kinetic heterogeneity becomes less.



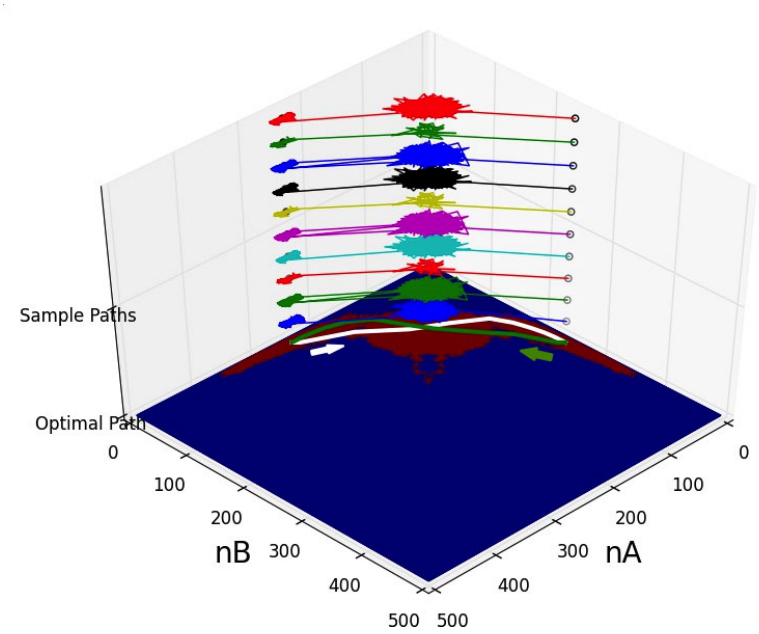
(a) FPT distribution at $\omega = 100$ (b) FPT distribution at $\omega = 0.1$ (c) FPT distribution at $\omega = 0.001$

Figure 8: FPT distribution from slow regulation non-adiabaticity to fast regulation adiabaticity

4.4.3 Cancer Heterogeneity from Kinetic Paths, Energy Cost and Stability

We can see the effects of heterogeneity on the kinetic paths from normal state to cancer state. We show the optimal kinetic paths when the adiabaticity parameter ω of relative binding/unbinding to synthesis and degradation is large (Fig. 9). We can see there is almost a unique path from normal state to cancer state through the intermediate premalignant state. The fluctuations around the optimal paths are small and the optimal path tube is narrow. However, when the adiabaticity parameter ω of relative binding/unbinding to synthesis and degradation is small (Fig. 10), the optimal path from normal state to cancer state deviates from the adiabatic optimal path. This is due to the emergence of more local state basins of attractions, leading to the shifts for the optimal paths. Furthermore, we see the fluctuations around this optimal path is relatively large compared with the adiabatic path, resulting a wider optimal path tube from normal to cancer state. Again, this is due to the presence of the multiple local state basins. Notice the discussion on the path fluctuations here in relation to the heterogeneity is consistent with the discussion on the kinetic hetero-

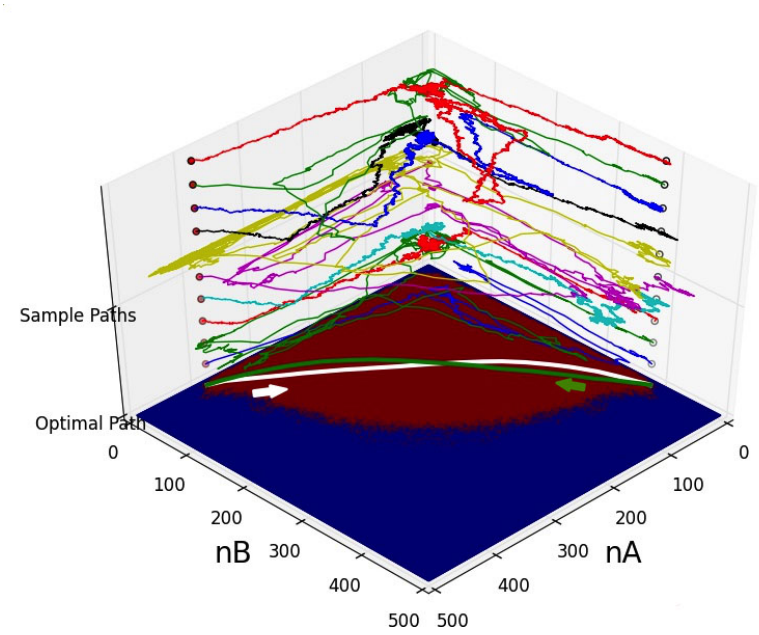
geneity (Fig. 8) where wider distribution of kinetics implies multiple paths from normal state to cancer state.



(a) $\omega = 100$ Adiabatic Approximation

Figure 9: Compare optimal path(white and green) at adiabatic limit under adiabatic approximation and real paths. Colored paths are real paths generated by Gillespie simulation

The cancer heterogeneity can also be reflected by the energy cost per gene switching. Being irreversible dynamic system, there is energy or heat dissipation measured by the entropy production associated with this irreversibility.[19, 23, 108] Fig. 10 shows the energy cost per gene switching from non-adiabaticity of slow regulation compared to protein synthesis/degradation to adiabaticity of fast regulation compared to protein synthesis/degradation. We see the energy cost per gene switching increases monotonically with respect to the slower regulation compared to the protein synthesis/degradation (decrease in ω). The increase of the energy cost is due to the heterogeneity in non-adiabatic slow regulation regime from



(a) $\omega = 0.1$ x-c Approximation

Figure 10: Compare optimal path(white and green) at moderate non-adiabatic region under unified landscape approximation and real paths. Colored paths are real paths generated by Gillespie simulation

the emergence of multiple state basins of attractions. Gene switchings cost more energy because of the number of state flipping is more.

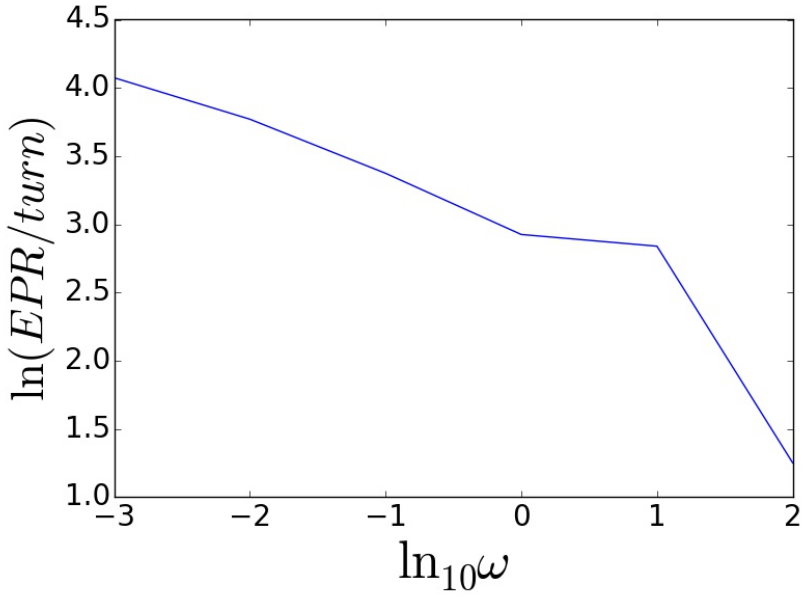


Figure 11: Energy cost per turnover from slow regulation non-adiabaticity to fast regulation adiabaticity

The heterogeneity can also be seen from the stability explorations. The stability of the normal and cancer state can be quantitatively measured by the kinetic time from one state to another. If the kinetic transition time from normal(cancer) state to cancer(normal) state is long, then the stability of normal(cancer) state is high. This is because the normal state can then be maintained for a long (short) time. In Fig. 11, we see the mean first passage time from normal to cancer state with respect to the adiabaticity parameter ω of gene regulation versus protein synthesis/degradation. The kinetics shows a non-monotonic behavior. At very low ω of extreme non-adiabatic regime with slow regulation, the rate limiting step is the gene switching. Increasing the regulation speed leads to the faster switching

and therefore faster kinetics. On the other hand, as the regulation time scale becomes faster than the protein synthesis and degradation, there is no sufficient time for protein copy numbers to reach to the transcription level corresponding to the gene state for each gene switching. Therefore, the rate limiting step is no longer gene switching anymore. In fact, the adiabatic barrier from one transcription (concentration) level to another is determined by the protein synthesis/degradation averaged over the rapid gene switchings. The faster the regulation speed or gene switchings, the harder for the gene transcription in terms of protein concentration level to catch for the specific gene state, therefore the kinetics becomes longer. Therefore both faster and slower regulation or switching give slower kinetics, while in the moderate regime of ω of gene regulation or switching relative to protein synthesis/degration, there is an optimal kinetic speed from normal to cancer state. Therefore, from the kinetics, we can see the stability is high at large ω (adiabatic regime). This is due to the limited state basins with significant depths. For smaller ω (non-adiabatic regime), the faster kinetics emerges. This implies less stability. The lower stability is from the the emergence of the multiple states with shallower basins. The very small ω of extreme non-adiabatic regime only has a moderate increase of the kinetic time scale and therefore slightly higher stability relative to the optimal one, in contrast with the adiabatic case for much longer duration and higher stability. As we can see the lower stability is another reflection of the underlying heterogeneity.

4.5 Conclusion

In this study, we use a core cancer gene regulatory motif to study the possible source of cancer heterogeneity. Normal state and cancer state, as well as intermediate states emerge as possible phenotype alternations

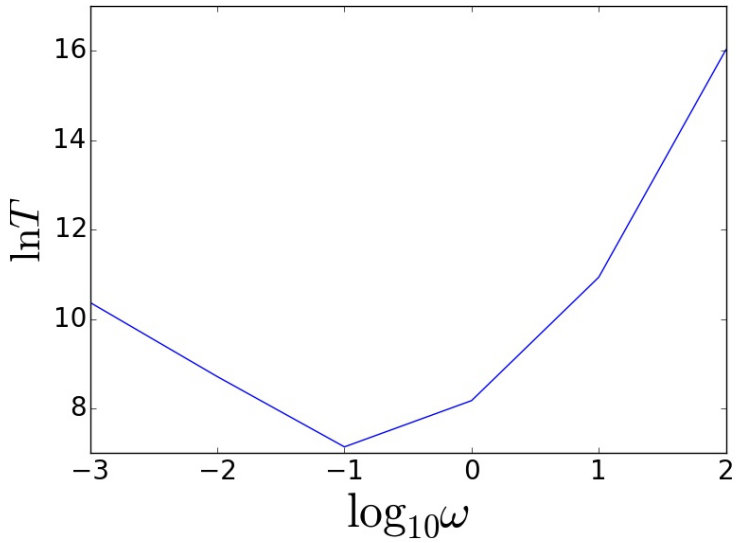


Figure 12: MFPT from slow regulation non-adiabaticity to fast regulation adiabaticity

in both adiabatic fast and non-adiabatic slow regulation regime. Slow regulations can come from epigentics of DNA methylation and histone remodification which lead to weaker coupling among genes. As a result, more steady states corresponding to more phenotype manifestations emerge. The cancer heterogeneity is reflected from the emergence of more phenotypic states, the larger transcription level concentration fluctuations, wider kinetic distributions and multiplicity of kinetic paths from normal to cancer state, more energy cost per gene switching, and weaker stability. The relationship between non-adiabatic slow regulation dynamics and epigenetic heterogeneity in cancer gene networks calls attention for further study.

Chapter 5

Conclusion

In this dissertation, we first give a brief introduction to gene network and its properties. Gene network is a mathematical model for the mechanism that controls gene expression through regulatory interactions. Each node of the network being a protein, RNA or protein complexes that is the transcription/translation products of a genes. The links or edges between nodes are the activating or repressive regulatory reactions that can be direct or indirect.

The biological properties of the cell is largely determined by the expression level of the nodes in the gene network. On the other hand, the transcription/translation products also participate in the regulatory interactions so that it is a network with feedback and network can respond to environment. A typical regulatory mechanism is through binding of a regulatory protein to the promoter site of the target gene and as a result activate or repress the gene. This is essential in biology. It is due to the feedback structure cells can respond to environment changes, adapt to changes and evolve. Gene network is a highly non-equilibrium system with frequent matter, energy and information exchange with outer environment. As a result, detailed balance is broken even at steady state after long time evolution. The level of non-equilibriumnes can be quantified with non-zero steady state flux and entropy production. Gene network is

also a stochastic system with intrinsic noise from biological chemical reactions involved and extrinsic noise from environment. Stochastic system is usually modeled with stochastic differential equation (SDE) which describes the stochastic trajectory of the system, and Fokker Planck Equation (FPE) which describes the evolution of the probability distribution. The dynamics is usually assumed to be Markovian and fluctuation is assumed to be space and time uncorrelated.

In the second chapter, we give a brief review of landscape and flux theory. Landscape theory visualizes system undergoes stochastic dynamics as charged particle moving on the potential energy surface. Landscape picture is useful in visualizing global distribution of stable biological states and quantifying global dynamical properties. We briefly summarize the relationship between stochastic Langevin/Fokker Planck system and Lagrangian system. We see in the sense of path integral, effective Lagrangian and corresponding path integral formalism can be constructed and the transition/escaping probability can be represented as path integral over all possible paths connecting initial and final states. Transition rate is recognized in this formalism. We further show in practice how to calculate transition rate and dominant path.

In the third chapter, we discuss in detail adiabatic and non-adiabatic dynamics of gene network. The time scale hierarchy of protein synthesis/degradation and regulation processes introduces different dynamics at adiabatic regime and non-adiabatic regime. Conventional study is concentrated in adiabatic regime where adiabatic approximation is valid and problem is greatly simplified. However, non-adiabatic dynamics is important in prokaryotic gene network and can have important biological implications. We show that by using spinor representation and path integral method we are able to map the 2^N dimensional gene network dynamics that is governed by master equation into a Langevin system in $2N$

extended space. This greatly reduces the complexity of the system and makes numerical study of large network possible. More importantly, it provides an unified landscape picture in the extended space. This provides a unified theoretical framework and is helpful in studying global and thermodynamical properties.

In the fourth chapter, we study the relationship between cancer heterogeneity and non-adiabatic network dynamics. It is now understood cancer is controlled by gene network rather than single gene mutation. We use a self-activating and mutual-repressing two gene motif as a simplified core cancer network. At all adiabaticity regime we see normal and cancer states. At moderate non-adiabatic regime we see in addition to normal and cancer states there's premalignant state emerges and multiple phenotypical variations of cancer state, a clear sign of heterogeneity. We further quantify heterogeneity by studying global stability that is measured by Fano factor. The distribution of first passage time, the mean first passage time at different adiabaticity, and diversity of transition paths all suggest moderate non-adiabatic regime has more heterogeneity and this heterogeneity is mainly due to the non-adiabatic dynamics of cancer network.

We get to the conclusion that non-adiabatic dynamics is essential to eukaryotic gene network evolution. It can have vast implication in real biological systems. We also have investigated a 6-gene core cancer network. Since this work is not completed yet we didn't include it in the dissertation. For the result we got so far, non-adiabatic dynamics in this network is able to explain stem cell, cancer, cancer stem cell states development, as well as cancer heterogeneity and cancer cancer metastasis. We hope to finish this project soon. In the future, since the unified landscape picture for non-adiabaric gene network dynamics is constructed and complexity of numerical simulation is greatly reduced, we hope further investigation into non-adiabatic dynamics of gene network can be done.

References

[1] Tong Ihn Lee, Nicola J Rinaldi, François Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, et al. Transcriptional regulatory networks in saccharomyces cerevisiae. *science*, 298(5594): 799–804, 2002.

[2] M Madan Babu, Nicholas M Luscombe, L Aravind, Mark Gerstein, and Sarah A Teichmann. Structure and evolution of transcriptional regulatory networks. *Current opinion in structural biology*, 14(3):283–291, 2004.

[3] Blaz Zupan, Janez Demšar, Ivan Bratko, Peter Juvan, John Halter, Adam Kuspa, and Gad Shaulsky. Genepath: a system for automated construction of genetic networks from mutant data. *Bioinformatics*, 19(3):383, 2003.

[4] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64–68, 2002.

[5] Jeff Hasty, Farren Isaacs, Milos Dolnik, David McMillen, and James J Collins. Designer gene networks: Towards fundamental cellular control. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 11 (1):207–220, 2001.

[6] N Barkal and Stan Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[7] George Von Dassow, Eli Meir, Edwin M Munro, and Garrett M Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

[8] Günter P Wagner and Jianzhi Zhang. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics*, 12(3):204–213, 2011.

[9] Eric H Davidson. *The regulatory genome: gene regulatory networks in development and evolution*. Academic Press, 2010.

[10] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126(4):663–676, 2006.

[11] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872, 2007.

[12] Anne BC Cherry and George Q Daley. Reprogramming cellular identity for regenerative medicine. *Cell*, 148(6):1110–1122, 2012.

[13] Chunhe Li and Jin Wang. Quantifying the underlying landscape and paths of cancer. *Journal of The Royal Society Interface*, 11(100): 20140774, 2014.

[14] Mingyang Lu, Kumar Jolly, Mohit, and Eshel Ben-Jacob. Toward decoding the principles of cancer metastasis circuits. *Cancer research*, 74(17):4574–4587, 2014.

[15] Robert A Gatenby and Thomas L Vincent. An evolutionary model of carcinogenesis. *Cancer Research*, 63(19):6212–6220, 2003.

[16] Pau Creixell, Erwin M Schoof, Janine T Erler, and Rune Linding. Navigating cancer network attractors for tumor-specific therapy. *Nature biotechnology*, 30(9):842–848, 2012.

[17] Chunhe Li and Jin Wang. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer research*, pages canres–0079, 2015.

[18] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403 (6767):339–342, 2000.

[19] Haidong Feng, Bo Han, and Jin Wang. Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *The Journal of Physical Chemistry B*, 115(5):1254–1261, 2010.

[20] Haidong Feng and Jin Wang. A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Scientific reports*, 2, 2012.

[21] Dominique Chu, Nicolae Radu Zabet, and Boris Mitavskiy. Models of transcription factor binding: sensitivity of activation functions to model assumptions. *Journal of Theoretical Biology*, 257(3):419–429, 2009.

[22] Liufang Xu, Hualin Shi, Haidong Feng, and Jin Wang. The energy pump and the origin of the non-equilibrium flux of the dynamical systems and the networks. *The Journal of chemical physics*, 136(16): 165102, 2012.

[23] Jin Wang, Li Xu, and Erkang Wang. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation,

and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271–12276, 2008.

[24] Jin Wang, Kun Zhang, and Erkwang Wang. Kinetic paths, time scale, and underlying landscapes: A path integral framework to study global natures of nonequilibrium systems and networks. *The Journal of chemical physics*, 133(12):125103, 2010.

[25] Xiaodong Cai and Xiaodong Wang. Stochastic modeling and simulation of gene networks–a review of the state-of-the-art research on stochastic simulations. 2007.

[26] Harley H McAdams and Adam Arkin. Itsa noisy business! genetic regulation at the nanomolar scale. *Trends in genetics*, 15(2):65–69, 1999.

[27] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94 (3):814–819, 1997.

[28] Jonathan M Raser and Erin K O'Shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005.

[29] Long Cai, Nir Friedman, and X Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440 (7082):358–362, 2006.

[30] F Nina and W Fontana. Genetic networks: small numbers of big molecules. *Science*, 297:1129–1131, 2002.

[31] Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415–418, 2004.

[32] Paul Smolen, Douglas A Baxter, and John H Byrne. Modeling transcriptional control in gene networksmethods, recent results, and future directions. *Bulletin of mathematical biology*, 62(2):247–292, 2000.

[33] Hidde De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, 9(1):67–103, 2002.

[34] E Atlee Jackson. *Perspectives of nonlinear dynamics*, volume 1. CUP Archive, 1992.

[35] Adam Arkin, John Ross, and Harley H McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

[36] Daniel T Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[37] M Andrecut, D Cloud, and SA Kauffman. Monte carlo simulation of a simple gene network yields new evolutionary insights. *Journal of theoretical biology*, 250(3):468–474, 2008.

[38] JEM Hornos, D Schultz, GCP Innocentini, JAMW Wang, AM Walczak, JN Onuchic, and PG Wolynes. Self-regulating gene: an exact solution. *Physical Review E*, 72(5):051907, 2005.

[39] Daniel Schultz, José N Onuchic, and Peter G Wolynes. Understanding stochastic simulations of the smallest genetic networks. *The Journal of chemical physics*, 126(24):245102, 2007.

[40] Haidong Feng, Bo Han, and Jin Wang. Landscape and global stability of nonadiabatic and adiabatic oscillations in a gene network. *Biophysical journal*, 102(5):1001–1010, 2012.

[41] Chunhe Li and Jin Wang. Quantifying waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *Journal of The Royal Society Interface*, 10(89):20130787, 2013.

[42] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. Stochastic mrna synthesis in mammalian cells. *PLoS Biol*, 4(10):e309, 2006.

[43] Masaki Sasai, Yudai Kawabata, Koh Makishi, Kazuhito Itoh, and Tomoki P Terada. Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLoS computational biology*, 9(12):e1003380, 2013.

[44] Aleksandra M Walczak, José N Onuchic, and Peter G Wolynes. Absolute rate theories of epigenetic stability. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18926–18931, 2005.

[45] RA Marcus. Chemical and electrochemical electron-transfer theory. *Annual Review of Physical Chemistry*, 15(1):155–196, 1964.

[46] John D Morgan and Peter G Wolynes. Adiabaticity of electron transfer at an electrode. *Journal of Physical Chemistry*, 91(4):874–883, 1987.

[47] Hu Gang. Stochastic forces and nonlinear systems. *Shanghai: Shanghai Scientific and Technological Education Publishing House*, 1994.

[48] JL Garcia-Palacios. Introduction to the theory of stochastic processes and brownian motion problems. *arXiv preprint cond-mat/0701242*, 2007.

[49] Hannes Risken. *Fokker-Planck Equation*. Springer, 1984.

[50] Peter G Wolynes, Jose N Onuchic, and D Thirumalai. Navigating the folding routes. *SCIENCE-NEW YORK THEN WASHINGTON-*, pages 1619–1619, 1995.

[51] Saul Lapidus, Bo Han, and Jin Wang. Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of mapk signal transduction. *Proceedings of the National Academy of Sciences*, 105(16):6039–6044, 2008.

[52] Jin Wang, Li Xu, Erkang Wang, and Sui Huang. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical journal*, 99(1):29–39, 2010.

[53] Jin Wang, Kun Zhang, Li Xu, and Erkang Wang. Quantifying the waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, 108 (20):8257–8262, 2011.

[54] Feng Zhang, Li Xu, Kun Zhang, Erkang Wang, and Jin Wang. The potential and flux landscape theory of evolution. *The Journal of chemical physics*, 137(6):065102, 2012.

[55] P Ao. Potential in stochastic differential equations: novel construction. *Journal of physics A: mathematical and general*, 37(3):L25, 2004.

[56] Adamo R Petosa, Deb P Jaisi, Ivan R Quevedo, Menachem Elimelech, and Nathalie Tufenkji. Aggregation and deposition of engineered nanomaterials in aquatic environments: role of physicochemical interactions. *Environmental science & technology*, 44(17): 6532–6549, 2010.

[57] Joseph Xu Zhou, MDS Aliyu, Erik Aurell, and Sui Huang. Quasi-potential landscape in complex multi-stable systems. *Journal of The Royal Society Interface*, 9(77):3539–3553, 2012.

[58] Andrzej M Kierzek. Stocks: Stochastic kinetic simulations of biochemical systems with gillespie algorithm. *Bioinformatics*, 18(3):470–481, 2002.

[59] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

[60] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[61] L Peliti. Path integral approach to birth-death processes on a lattice. *Journal de Physique*, 46(9):1469–1483, 1985.

[62] Robert S Maier and Daniel L Stein. Escape problem for irreversible systems. *Physical Review E*, 48(2):931, 1993.

[63] Katharine LC Hunt and John Ross. Path integral solutions of stochastic equations for nonlinear irreversible processes: the uniqueness of the thermodynamic lagrangian. *The Journal of Chemical Physics*, 75(2):976–984, 1981.

[64] Sidney Coleman. Fate of the false vacuum: Semiclassical theory. *Physical Review D*, 15(10):2929, 1977.

[65] B Caroli, C Caroli, and B Roulet. Diffusion in a bistable potential: The functional integral approach. *Journal of Statistical Physics*, 26(1): 83–111, 1981.

[66] Kun Zhang, Masaki Sasai, and Jin Wang. Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. *Proceedings of the National Academy of Sciences*, 110 (37):14930–14935, 2013.

[67] Haidong Feng, Kun Zhang, and Jin Wang. Non-equilibrium transition state rate theory. *Chemical Science*, 5(10):3761–3769, 2014.

[68] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001. doi: 10.1073/pnas.151588598.

[69] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[70] Jin Wang, Li Xu, ErKang Wang, and Sui Huang. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical Journal*, 99:29–39, 2008. doi: 10.1016/j.bpj.2010.03.058.

[71] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes. Self-regulating gene: An exact solution. *Phys. Rev. E*, 72:051907, Nov 2005. doi: 10.1103/PhysRevE.72.051907.

[72] Masaki Sasai and Peter G Wolynes. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences*, 100(5):2374–2379, 2003.

[73] Frank Jülicher, Armand Ajdari, and Jacques Prost. Modeling molecular motors. *Rev. Mod. Phys.*, 69:1269–1282, Oct 1997. doi: 10.1103/RevModPhys.69.1269.

[74] Hong Qian. The mathematical theory of molecular motor movement and chemomechanical energy transduction. *Journal of Mathematical Chemistry*, 27(3):219–234, 2000. ISSN 0259-9791. doi: 10.1023/A:1026428320489. URL http://dx.doi.org/10.1023/A%3A1026428320489.

[75] Qing-Miao Nie, Akio Togashi, Takeshi N Sasaki, Mitsunori Takano, Masaki Sasai, and Tomoki P Terada. Coupling of lever arm swing and biased brownian motion in actomyosin. *PLoS Comput. Biol.*, 10 (4):e1003552, 2014. doi: 10.1371/journal.pcbi.1003552.

[76] Gaper Tkaik and Aleksandra M Walczak. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, 23:153102, 2011. doi: 10.1088/0953-8984/23/15/153102.

[77] Thomas B Kepler and Timothy C Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical journal*, 81(6):3116–3136, 2001.

[78] Richard Phillips Feynman and Albert R Hibbs. *Quantum mechanics and path integrals*, volume 2. McGraw-Hill New York, 1965.

[79] Bin Zhang and Peter G Wolynes. Stem cell differentiation as a many-body problem. *Proceedings of the National Academy of Sciences*, 111(28): 10185–10190, 2014.

[80] E de Wit, Britta A M Bouwman, Yun Zhu, Petra Klous, Erik Splinter, Marjon J A M Verstegen, Peter H L Krijger, Nicola Festuccia, E P Nora, Maaike Welling, Edith Heard, Niels Geijsen, Raymond A Poot, Ian Chambers, and Wouter de Laat. *Nature*, 501:227, 2012.

[81] Hang Zhang, Xiao-Jun Tian, Abhishek Mukhopadhyay, K. S. Kim, and Jianhua Xing. Statistical mechanics model for the dynamics of collective epigenetic histone modification. *Phys. Rev. Lett.*, 112:068101, Feb 2014. doi: 10.1103/PhysRevLett.112.068101. URL `http://link.aps.org/doi/10.1103/PhysRevLett.112.068101`.

[82] S S Ashwin and Masaki Sasai. *http://arxiv.org/abs/1410.2337*, 2014.

[83] K Sneppen and I B Dodd. *PLoS Comput. Biol.*, 8:e1002643, 2012.

[84] Haidong Feng and Jin Wang. Potential and flux decomposition for dynamical systems and non-equilibrium thermodynamics: Curvature, gauge field, and generalized fluctuation-dissipation theorem. *The Journal of Chemical Physics*, 135(23):234511, 2011. doi: http://dx.doi.org/10.1063/1.3669448.

[85] Haidong Feng and Jin Wang. Potential and flux decomposition for dynamical systems and non-equilibrium thermodynamics: Curvature, gauge field, and generalized fluctuation-dissipation theorem. *The Journal of chemical physics*, 135(23):234511, 2011.

[86] Andriy Marusyk, Vanessa Almendro, and Kornelia Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer*, 12(5):323–334, 2012.

[87] Tannishtha Reya, Sean J Morrison, Michael F Clarke, and Irving L Weissman. Stem cells, cancer, and cancer stem cells. *nature*, 414 (6859):105–111, 2001.

[88] John E. Dick. Stem cell concepts renew cancer research. *Blood*, 112(13):4793–4807, 2008. ISSN 0006-4971. doi: 10.1182/blood-2008-08-077941.

[89] Mark Shackleton, Elsa Quintana, Eric R. Fearon, and Sean J. Morrison. Heterogeneity in cancer: Cancer stem cells versus clonal evolution. *Cell*, 138(5):822 – 829, 2009. ISSN 0092-8674. doi: http://dx.doi.org/10.1016/j.cell.2009.08.017. URL http://www.sciencedirect.com/science/article/pii/S0092867409010307.

[90] Muhammad Al-Hajj, Max S. Wicha, Adalberto Benito-Hernandez, Sean J. Morrison, and Michael F. Clarke. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, 100(7):3983–3988, 2003. doi: 10.1073/pnas.0530291100. URL http://www.pnas.org/content/100/7/3983.abstract.

[91] Sheila K Singh, Cynthia Hawkins, Ian D Clarke, Jeremy A Squire, Jane Bayani, Takuichiro Hide, R Mark Henkelman, Michael D Cusimano, and Peter B Dirks. Identification of human brain tumour initiating cells. *nature*, 432(7015):396–401, 2004.

[92] Piero Dalerba, Scott J. Dylla, In-Kyung Park, Rui Liu, Xinhao Wang, Robert W. Cho, Timothy Hoey, Austin Gurney, Emina H. Huang, Diane M. Simeone, Andrew A. Shelton, Giorgio Parmiani, Chiara Castelli, and Michael F. Clarke. Phenotypic characterization of human colorectal cancer stem cells. *Proceedings of the National Academy of Sciences*, 104(24):10158–10163, 2007. doi: 10.1073/pnas.0703478104. URL http://www.pnas.org/content/104/24/10158.abstract.

[93] Andrei V Krivtsov, David Twomey, Zhaohui Feng, Matthew C Stubbs, Yingzi Wang, Joerg Faber, Jason E Levine, Jing Wang, William C Hahn, D Gary Gilliland, et al. Transformation from committed progenitor to leukaemia stem cell initiated by mll–af9. *Nature*, 442(7104):818–822, 2006.

[94] Sui Huang, Ingemar Ernberg, and Stuart Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. In *Seminars in cell & developmental biology*, volume 20, pages 869–876. Elsevier, 2009.

[95] Mads Kaern, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[96] Jiyoung Lee, Jinho Lee, Kevin S Farquhar, Jieun Yun, Casey A Frankenberger, Elena Bevilacqua, Kam Yeung, Eun-Jin Kim, Gábor Balázsi, and Marsha Rich Rosner. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proceedings of the National Academy of Sciences*, 111(3): E364–E373, 2014.

[97] Mingyang Lu, Mohit Kumar Jolly, Ryan Gomoto, Bin Huang, Jos Onuchic, and Eshel Ben-Jacob. Tristability in cancer-associated microrna-tf chimera toggle switch. *The Journal of Physical Chemistry B*, 117(42):13164–13174, 2013.

[98] Matteo Parri and Paola Chiarugi. Rac and rho gtpases in cancer cell motility control. *Cell Commun Signal*, 8(23):10–1186, 2010.

[99] Dawn R Cochrane, Erin N Howe, Nicole S Spoelstra, and Jennifer K Richer. Loss of mir-200c: a marker of aggressiveness and chemoresistance in female reproductive cancers. *Journal of oncology*, 2010, 2009.

[100] Raúl Guantes and Juan F Poyatos. Multistable decision switches for flexible control of epigenetic differentiation. *PLoS Comput Biol*, 4(11): e1000235–e1000235, 2008.

[101] Mingyang Lu, José Onuchic, and Eshel Ben-Jacob. Construction of an effective landscape for multistate genetic switches. *Physical review letters*, 113(7):078102, 2014.

[102] SS Ashwin and Masaki Sasai. Epigenetic dynamics of cell reprogramming. *arXiv preprint arXiv:1410.2337*, 2014.

[103] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[104] Roberto Olender and Ron Elber. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *The Journal of chemical physics*, 105(20):9299–9315, 1996.

[105] P. Faccioli, M. Sega, F. Pederiva, and H. Orland. Dominant pathways in protein folding. *Phys. Rev. Lett.*, 97:108101, Sep 2006. doi: 10.1103/PhysRevLett.97.108101. URL `http://link.aps.org/doi/10.1103/PhysRevLett.97.108101`.

[106] Siddhartha Jaiswal, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V Grauman, Brenton G Mar, R Coleman Lindsley, Craig H Mermel, Noel Burtt, Alejandro Chavez, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine*, 371(26):2488–2498, 2014.

[107] Giulio Genovese, Anna K Kähler, Robert E Handsaker, Johan Lindberg, Samuel A Rose, Samuel F Bakhoum, Kimberly Chambert, Eran Mick, Benjamin M Neale, Menachem Fromer, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood dna sequence. *New England Journal of Medicine*, 371(26):2477–2487, 2014.

[108] Hong Qian. Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. *Physical Review E*, 65(1):016102, 2001.