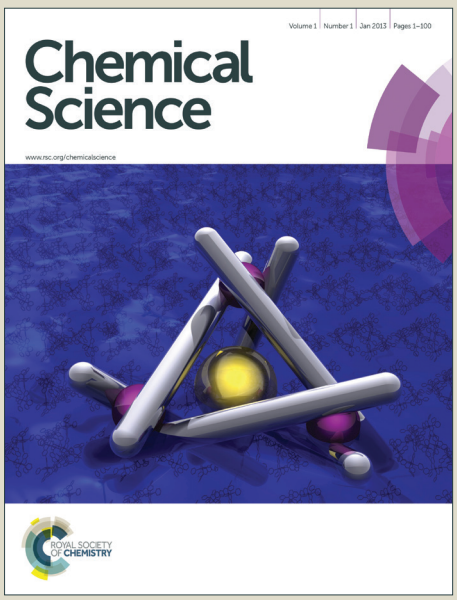
# Chemical Science

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemicalscience

Chemical Science
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

# Non-equilibrium Transition State Rate Theory†

**H. D. Feng[1], K. Zhang[2] and J. Wang[1,2,3,‡]**

***Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX***
***First published on the web Xth XXXXXXXXXX 200X***
**DOI: 10.1039/b000000x**

Transition state or Kramers' rate theory has been used to quantify the kinetic speed of many chemical, physical and biological equilibrium processes successfully. For non-equilibrium systems, the analytical quantification of the kinetic rate is still challenging. We developed a new transition state or Kramers' rate theory for general non-equilibrium stochastic systems with finite fluctuations. We illustrated that the non-equilibrium rate is mainly determined by the exponential factor as the weight action measured from the basin of attraction to the "saddle" or more accurately "global maximum" point on the optimal path rather than the saddle point of the underlying landscape as in the conventional transition state or Kramers' rate formula for equilibrium systems. Furthermore, the pre-factor of the non-equilibrium rate is determined by the fluctuations around the basin of attraction and "saddle" point along the optimal paths. We apply our theory for non-equilibrium rate to fate decisions in stem cell differentiation. The dominant kinetic paths between stem and differentiated cell basins are irreversible and do not follow the gradient path along the landscape. This reflects that the dynamics of non-equilibrium systems is not only determined by the landscape gradient but also the curl flux, suggesting experiments to test theoretical predictions. We calculated the transition rate between cell fates. The predictions are in good agreements with stochastic simulations. Our general rate and path formula can be applied to other non-equilibrium systems.

## 1 Introduction

For complex chemical and biological systems, identifying the most important dynamic flow and estimating the transition rates from one stable state in a basin of attraction defining an equilibrium or nonequilibrium chemical state under fluctuations, to another is crucial in understanding the underlying kinetic mechanisms and global robustness[1,2]. Furthermore, the driving force of many dynamical systems in chemical and physical world can not be written in terms of the pure gradient of a potential, which is closely linked to the underlying non-equilibrium natures[3]. For example, the setups for the normal bulk enzyme kinetic experiments are sometimes in non-equilibrium conditions such as constant flow. In single molecule enzymatic experiments, the substrate concentration is high and can be thought of not changing significantly. This often creates non-equilibrium yet steady conditions for enzyme kinetic measurements[4–11].

For the equilibrium systems, the global stability and robustness of a complex stochastic system can be quantitatively studied if the underlying potential landscape is known a priori. For instance, the dynamics and the dominant (the most probable) kinetic transition paths between different states follow the gradient ascending or descending on the potential landscape. Furthermore, the famous transition state or Kramers' rate formula for kinetic speed is determined by the barrier between the basins of attraction (the barrier height is determined by the difference in energy between the stable fixed point and the saddle point on the underlying potential landscape) and the fluctuations around one basin and the saddle point between the basins of attractions. This was proposed by Eyring in chemistry and Kramers in physics on thermally activated barrier crossing more than 70 years ago[12,13]. It provides a good analytical approximated formula of the transition rate from one attractor to another for equilibrium systems with small fluctuations. However, for general non-equilibrium dynamical systems, such as gene regulatory networks or enzymatic reactions, transition state rate formula often fails because the dominant dynamic paths are not reversible and do not follow the gradient path of the underlying non-equilibrium potential landscape[14].

For non-equilibrium stochastic processes with constant diffusions, several approaches have been proposed to identify the dominant (optimal) transition paths between arbitrary states, especially between stable states[14–22]. Such formalism has wide applications ranging from equilibrium dynamics such

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

[1]*Department of Chemistry, Physics and Applied Mathematics, State University of New York at Stony Brook, Stony Brook, NY, 11794 USA*
[2]*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, 130022, P. R. China*
[3]*College of Physics and State Key laboratory of Super Hard Materials, Jilin University, Changchun, Jilin, 130021, P. R. China.*
[‡]*E-mail: jin.wang.1@stonybrook.edu; Tel: +1-631-632-1185; Fax: +1-631-632-7960*

as protein folding to non-equilibrium systems such as gene regulation networks[14,20,23–26]. However, many chemical and biological systems have finite, non-constant, local diffusion depending on the underlying variables, for example, concentrations. Such location dependent diffusion coefficients might have significant impact on kinetic paths as well as transition rates. Therefore, a complete theoretical formalism of non-equilibrium paths accounting for the non-constant diffusion will be natural and necessary.

More importantly, the equilibrium transition state rate can be estimated by the path integral formalism[27]. However, an analytical formalism of the transition state rate in non-equilibrium systems, which measures the capability of communicating between stable states and therefore the global robustness, is still challenging. It was argued that, in the zero noise limit, the dominant path will go through the saddle point between the basins of attraction and an analytic approximation of the escape rate from the basin can be derived[15,17,18]. However, in general non-equilibrium systems, small but finite fluctuations often emerge, and the dominant kinetic paths do not necessary go through the saddle points[14].

In this work, we developed a new analytical transition state theory for kinetic rate of general non-equilibrium systems. In this formula (theory), (i) we first obtain the most probable path according to the path integral by minimizing the action. Here the starting point and the ending point for the path integral are the two stable fixed points $S$ and $S'$. (ii) Because of the non-zero flux, this most probable path will not follow the gradient path of the landscape. In addition, under finite fluctuations this most probable path may not even go through the original saddle point $\hat{S}$. (iii) Then, on the most probable path, we search for $\hat{S}'$, the new "saddle" or more accurately "global maximum along the dominant path". For general non-equilibrium systems under finite noise, $\hat{S}'$ will not likely to be at the original saddle point $\hat{S}$ of the driving force. (iv) The action of the path integral from $S$ to $S'$ obtained in (i), which is the minimal among all paths, is smaller than the action through the original saddle $\hat{S}$ or the action along the gradient path of landscape. This action calculated along the dominant path from the stable fixed point to the "global maximum along the dominant path" will give the exponential term in our new transition state rate theory. As the comparison, in conventional transition state theory for equilibrium systems, the kinetic rate is only determined by the barrier from the saddle point between the basins of attraction on the underlying landscape. Furthermore, the pre-factor part reflects the fluctuations around the basin and the local curvature around the new saddle along the optimal path.

As an application, we will study an important example of cell developmental circuit composed of a pair of self-activating and mutually inhibiting genes. In various tissues, this gene regulatory motif has been considered to control bi-

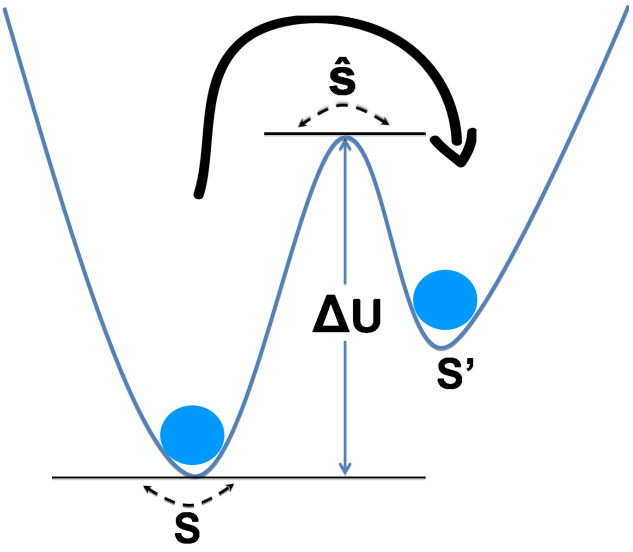



**Fig. 1** (Color online) The potential barrier $\Delta U$ for calculating the transition state or Kramers' escaping rate. The basins of attractions are localized at $S$ and $S'$. $\hat{S}$ is the saddle point.

nary cell fate decisions in pluri/multipotent stem cells[28–34]. For example, the multipotent common myeloic progenitor cell (CMP) faces the binary cell fate decision between the myeloid and the erythroid fate. Such fate commitments are determined by transcription factors (TF), PU.1, and GATA1. The relative expression levels A (PU.1) and B (GATA1) of these two reciprocal TFs can promote the decision towards either lineage[31,33]. For this system, we show that kinetic dominant paths for the differentiation and reprogramming are irreversible and do not pass through the saddle points of the underlying potential landscape. Using our newly developed transition state theory for kinetic rate of non-equilibrium systems, we also estimate the transition rates of the differentiation from our new formula (theory), which agree with the stochastic simulation results within the same order of magnitude. This framework can be applied to other general non-equilibrium chemical systems such as enzymatic kinetics, whose reaction rate can be easily measured by enzyme assays[35], networks and dynamical systems.

## 2 Results and Discussions

### 2.1 Equilibrium Transition State Rate

In this section, we will first review the equilibrium transition state or Kramer's rate theory. The stochastic dynamics can be quantified in continuous spaces by the *Langevin* equations (in *Ito*'s form): $\dot{x}_\mu = F_\mu(\vec{x}) + \sum_a B^a_\mu(\vec{x})\xi^a(t)$, where $\vec{x}$ represents the dynamical variables of the system. $F_\mu(\vec{x})$ is the driving force. $\xi_a(t)$ represents the Gaussian distributed white

noise unit fluctuations. $B^q_\mu(\vec{x})$ represents the strength or magnitude of the variable dependent fluctuations: $\langle \xi^a(t)\xi^b(t')\rangle = \delta^{ab}\delta(t-t')$. In addition, rather than each individual trajectory, the corresponding probability $P(\vec{x},t)$ obeys the *Fokker-Planck equation*[36]:

$$\frac{dP}{dt} = \sum_\mu \partial_\mu(-F_\mu P) + \sum_{\mu,\nu}\frac{1}{2}\partial_\mu\partial_\nu(\varepsilon_{\mu\nu}P) \quad (1)$$

with the diffusion coefficient $\varepsilon_{\mu\nu}(\vec{x}) = \sum_{a,b} B^a_\mu(\vec{x})B^b_\mu(\vec{x})\delta^{ab}$. Here, we use the notation $\partial_\mu \equiv \frac{\partial}{\partial x_\mu}$. For convenience, we also use $P(\vec{x}) \equiv P(\vec{x},t)$ to represent the time dependent probability distribution and $P^{SS}(\vec{x})$ to indicate the time independent steady state probability distribution. When considering specifically the intrinsic noise from molecular number fluctuations, the resulting the *Fokker-Planck* equation, whose diffusion coefficients depend on the location $\vec{x}$ (concentrations), can be derived from the second order Taylor expansion of the underlying chemical master equations (CME) describing the intrinsic fluctuations[37].

The *Fokker-Planck equation* can be rewritten in the format of probability conservation where the local probability change is equal to the net in or out flux: $\frac{dP(\vec{x},t)}{dt} = -\partial\cdot\mathbf{j}$. The system is considered to be in detailed balance if the steady state flux: $F_\mu(\vec{x})P^{SS}(\vec{x}) - \sum_\nu \frac{1}{2}\partial_\nu[\varepsilon_{\mu\nu}(\vec{x})P^{SS}(\vec{x})] = j^{SS}_\mu(\vec{x})$ is zero: $\vec{j}^{SS} = 0$. In this case, the system is in equilibrium state. The equilibrium probability distribution is closely related to the underlying potential and the driving force is determined by the gradient of the equilibrium potential: $U = -lnP_{eq}$ and $F_\mu(\vec{x}) = -\frac{1}{2}\partial_\mu[U(\vec{x})] + \sum_\nu \frac{1}{2}\partial_\nu[\varepsilon_{\mu\nu}(\vec{x})]$. For general non-equilibrium systems without detailed balance, the flux is not zero, $\vec{j}^{SS} \neq 0$, the steady state flux is a divergence free vector with $\vec{\partial}\cdot\vec{j}^{SS} = 0$ reflecting its rotational curl nature. The flux quantifies the degree of how far the system is away from the equilibrium. For non-equilibrium dynamical systems, the dynamics determined by the driving force ($F_\mu(\vec{x}) = -\frac{1}{2}\partial_\mu[U(\vec{x})] + \sum_\nu \frac{1}{2}\partial_\nu[\varepsilon_{\mu\nu}(\vec{x})] + \vec{j}^{SS}/P_{ss}$ ) and the global nature are quantified by both the steady state probability distribution which defines the non-equilibrium landscape $U = -lnP_{ss}$ and the curl probability flux: $j^{SS}_\mu(\vec{x})$.

For one dimensional systems, which are integrable with detailed balance $j^{SS} = 0$ under nature boundary conditions, transition state theory for the escaping rate from one basin to another basin gives

$$r^{eq}_K = \frac{\sqrt{U''(S))|U''(\hat{S})|}}{2\pi}e^{-2[U(\hat{S})-U(S)]/\varepsilon} \quad (2)$$

Here, $U(x)$ is the potential function in the equilibrium system and the driving force is a gradient of the potential: $F(x) = -U'(x)$, as shown in Fig. 1. The attractor is located at $S$ and the saddle point of the potential landscape is at $\hat{S}$ where the barrier is. In the small noise limit $\varepsilon \to 0$, the transition state rate in equation (2) can also be rewritten as[38]

$$r^{eq}_K = (2\pi)^{-1}\sqrt{-\frac{dF}{dx}(S)\frac{dF}{dx}(\hat{S})}e^{-S^{DOM}_{HJ}} \quad (3)$$

in which $S^{DOM}_{HJ} = \int_S^{\hat{S}} p\cdot dx$ is the *HJ weight action* or work (Hamilton-Jacobi weight action) along the dominant path from $S$ to $\hat{S}$[14,27] where $p$ is the canonical momentum and $dx$ is the variable displacement of the system. The physical meaning is clear. The transition state rate for equilibrium process is determined by two factors. The dominant factor is determined by the exponential of the weight action. The other is the prefactor determined by the fluctuations around stable point and saddle or transition state of the underlying equilibrium potential landscape. The 1 dimensional transition state rate as in equation (2) can be generalized into the N dimensional form[13,39,40] for equilibrium systems. However, for general non-equilibrium systems, the driving force can not be written as a pure gradient and there is no well defined potential $U$ to give the driving force as a gradient of a potential, $F_\mu(\vec{x}) = -\partial_\mu U(\vec{x})$. In addition, without the detailed balance, the curl current flux $\vec{j}^{ss}$ is not zero, which can lead the transition path to deviate from the gradient one and the *HJ weight action* $S^{DOM}_{HJ}$ becomes path dependent[14]. The dominant or optimal paths may not pass through the saddle points or transition states. Therefore, new transition state rate for general non-equilibrium systems is needed. We have to specify the transition path as well as the complete form of the weight action in order to quantify the rates from transition states might be path dependent in contrast to the equilibrium case where they are fixed.

### 2.2 Exponential Factor of Non-equilibrium Transition State Rate

For a 2 (or N) dimensional non-equilibrium system, we need to develop a new transition state theory for kinetic rates beyond the equilibrium one dictated by equation (3). The generalized weight action for non-equilibrium systems can be derived as (**for details see supporting information**): $S = \int_{t_i}^{t_f} dt\mathscr{L}$ with the *Lagrangian*

$$\mathscr{L} = \sum_{\mu\nu}\frac{\varepsilon^{-1}_{\mu\nu}}{2}(\dot{x}_\mu - F_\mu)(\dot{x}_\nu - F_\nu) + \sum_{\mu\nu\chi}\frac{1}{2}\varepsilon_{\mu\chi}\partial_\chi(F_\nu\varepsilon^{-1}_{\mu\nu}) \quad (4)$$

In the zero fluctuation limit $\varepsilon \to 0$, the action $S = \int_{t_i}^{t_f} dt\mathscr{L}$ leads to the exponential part of the Freidlin-Wentzell's theory[15]. In addition, in the zero fluctuation limit, the ratio of $e^{-S_{l_1}}/e^{-S_{l_2}}$ between the two smooth paths $l_1$ and $l_2$ agrees with Onsager-Machlup function[16].

The optimal path which contributes most to the path weight by taking the functional variations of the weight action $S$ with

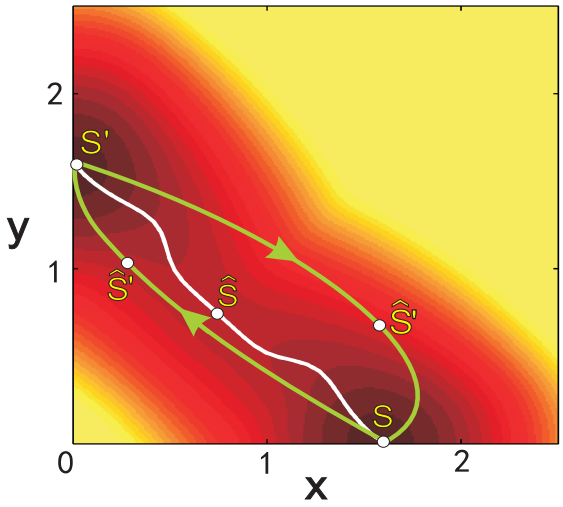



**Fig. 2** (Color online) 2D illustration of non-equilibrium landscape with the irreversible dominant transition paths between basins $S$ and $S'$ (green lines with arrows) and the gradient path (white line). Here, $\hat{S}$ is the saddle point and $\hat{S}'$ is the "global maximum along the dominant path".

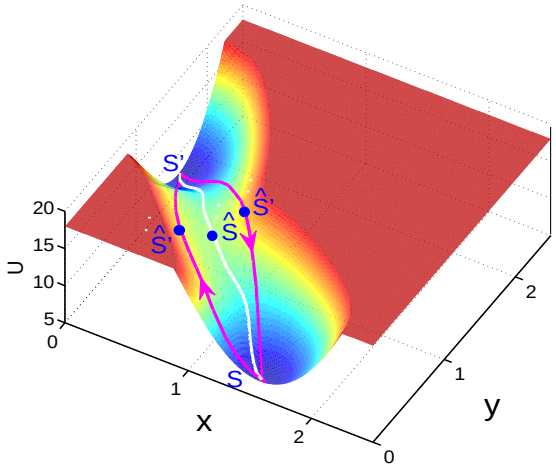



**Fig. 3** (Color online) 3D illustration of non-equilibrium landscape with the irreversible dominant transition paths between basins $S$ and $S'$ (purple lines with arrows) and the gradient path (white line)). Here, $\hat{S}$ is the saddle point and $\hat{S}'$ is the "global maximum along the dominant path".

respect to the paths $x_\mu(t)$'s, we obtain the equation of motion for the dominant path which satisfies the *Euler-Lagrangian* equation $\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{x}_\alpha} = \frac{\partial \mathscr{L}}{\partial x_\alpha}$. The dominant path approach gives the lowest order approximation of the full path integral weight action[14].

Instead of solving the *Euler-Lagrangian* equation of motion directly, the dominant kinetic path can also be evaluated by minimizing the weight action $S$ in path integral formalism. Define canonical momentum $p_\mu = \frac{\partial \mathscr{L}}{\partial \dot{x}_\mu} = \sum_\nu \varepsilon^{-1}_{\mu\nu}(\dot{x}_\nu - F_\nu)$, the total energy

$$-E = -H = \mathscr{L} - p_\mu \dot{x}_\mu \tag{5}$$

conserves along the dominant kinetic path. Then, the *HJ weight action*[41], which should be minimized to find the dominant path, can be written as (**for details see supporting information**):

$$S_{HJ}(x_i, x_f) = \int_{x_i}^{x_f} \sqrt{2(E - V_{eff})}dl - \int_{x_i}^{x_f} \sum_{\mu\nu} \varepsilon^{-1}_{\mu\nu} F_\nu dx_\mu \tag{6}$$

which is simplified to a line integral along the dominant path $dl = \sqrt{\sum_{\mu\nu} \varepsilon^{-1}_{\mu\nu} dx_\mu dx_\nu}$ in a curved space with distance measure $\varepsilon^{-1}_{\mu\nu}$ where the $\varepsilon_{\mu\nu}$ characterize the fluctuation or diffusion strengths.

### 2.3 New Transition State Rate for Non-equilibrium Systems

We found the forward and backward dominant paths (lines with arrows) are irreversible and do not go through the saddle point $\hat{S}$ on the gradient path (white lines) along the landscape. However, using the effective driving force $F^{eff}_\mu = \sum_\nu \varepsilon^{-1}_{\mu\nu} F_\nu$ in the 2nd term on the right side of equation (6), we can always find the "global maximum along the dominant path" $\hat{S}'$ with the component $F^{eff}_l(S')$ along the path is zero, as shown in in Fig. 2 (2D) and Fig. 3 (3D)! This is because $F^{eff}_l$ always changes its sign from the neighborhood of $S$ (pointing to $S$) to the neighborhood $S'$ (pointing to $S'$). Normally, there will be only one "global maximum along the dominant path" (or one new saddle point along the path), since the new saddle is between one basin or the other. Multiple new "saddles" along the path will introduce additional basin of attractions between them. In the extreme case when there are multiple new "saddles" along the path, we choose the last one before reaching the ending stable fixed point S as $\hat{S}'$. Therefore, by replacing the saddle point $\hat{S}$ for equilibrium systems by the "global maximum along the dominant path" $\hat{S}'$ for the non-equilibrium system, we can derive a new analytical transition state theory for kinetic rates of non-equilibrium systems (**details in supporting information**) as

$$r_K^{noneq} = (E\tau)^{-1} = \frac{\lambda_u(\hat{S}')}{2\pi}\sqrt{\frac{detM(S)}{|detM(\hat{S}')|}}e^{-S^{DOM}_{HJ}} \tag{7}$$

where the *HJ weight action* $S_{HJ}^{DOM} = \int_S^{\hat{S}'} p \cdot dx$ is integrated along the dominant path from the fixed point (basin) at $S$ to the "global maximum along the dominant path" at $\hat{S}'$.

For the pre-factor, we follow the similar derivation as the case of zero noise limit[18] (**details reviewed in supporting information**). Here, $\lambda_u(\hat{S}')$ is the positive eigenvalue of force matrix $F_{\mu,\nu}(\hat{S}') = \frac{\partial F_\mu}{\partial x_\nu}(\hat{S}')$ at the "global maximum along the dominant path" $\hat{S}'$, which represents the fluctuations along the dominant path at $\hat{S}'$. At the stable state $S$, we have the stationary solution for the *Fokker-Planck* equation and the matrix $M(S)$ satisfy the algebra equation (8) at the stable state $S$:

$$\sum_{\xi\chi} \varepsilon_{\xi\chi} M_{,\mu\xi} M_{,\nu\chi} + \sum_{\xi} M_{,\mu\xi} F_{\nu,\xi} + \sum_{\xi} M_{,\nu\xi} F_{\mu,\xi} = 0 \qquad (8)$$

At the "global maximum along the dominant path" $\hat{S}'$, since it is not a fixed point (force $F \neq 0$), we do not have a stationary solution for the *Fokker-Planck* equation at $\hat{S}'$ and the matrix $M(\hat{S}')$ satisfies the dynamic equation at $\hat{S}'$:

$$\frac{dM_{\mu\nu}(x)}{dt} = \frac{\partial^2 H}{\partial p_\xi \partial p_{\xi'}} M_{\mu\xi} M_{\nu\xi'} - \frac{\partial^2 H}{\partial x_\mu \partial x_\nu} - \frac{\partial^2 H}{\partial x_\nu \partial p_\xi} M_{\mu\xi} - \frac{\partial^2 H}{\partial x_\mu \partial p_\xi} M_{\nu\xi} \qquad (9)$$

The $\sqrt{\frac{detM(S)}{|detM(\hat{S}')|}}$ represents the ratio of the curvature around saddle along the dominant path and stable basin state ($detM(S)$ represents the second order fluctuations around stable basin state in all directions, while $detM(S')$ represents the second order fluctuations around the "saddle point" $\hat{S}'$ in all directions). In other words, $\lambda_u(\hat{S}')$ measures the frequency or fluctuations of the single unstable mode at the saddle-point $\hat{S}'$, $detM(S)$ measures the fluctuations in terms of frequencies of all stable modes at $S$, and $detM(S')$ measures the fluctuations in terms of frequencies of all stable modes and unstable modes at $\hat{S}'$. For the exponential factor, the weight action $S_{HJ}^{DOM}$ represents the weight action $S_{HJ}$, as defined in equation (6), calculated along the one dimensional dominant path $l$ from the stable basin $S$ to the "global maximum along the dominant path" $\hat{S}'$. In this rate expression for non-equilibrium dynamical systems, the major contribution comes from the exponential term with the weight action from stable basin state to the saddle point on the dominant path (based on the path integral formalism in a curved length space). While the non-exponential pre-factor gives the second order correction or fluctuations determined locally at the stable point $S$ and the "global maximum along the dominant path" $\hat{S}'$.

On the one hand, in conventional transition state theory for equilibrium systems, the kinetic rate is determined by the saddle point on the underlying landscape or more explicitly the potential barrier between the basins of attraction (potential difference between the saddle point and stable basin on the landscape). On the other hand, in our kinetic rate formula (theory) for non-equilibrium systems, the kinetic rate is determined by the weight action along the dominant path from the basin to the "global maximum along the dominant path" . We see the non-equilibrium 'saddle point' is path and directional dependent (the forward and backward paths do not share the same 'saddle point' as in the conventional equilibrium case as shown in Fig. 5). In addition, although equation (7) is derived in 2 dimensional space, it can be generalized to any dimension with the same final form.

### 2.4 Compare with Transition State Rate of Zero Noise approximation

With the zero fluctuation approximations[15,17,18], Transition State Rate can be written as

$$r_K^{noneq} = (E\tau)^{-1} = \frac{\lambda_u(\hat{S})}{2\pi} \sqrt{\frac{detM(S)}{|detM(\hat{S})|}} K(\hat{S}) e^{-\int_S^{\hat{S}} \Sigma_\mu p_\mu dx_\mu} \qquad (10)$$

Here, because the possibility of crossing the separatrix in either directions equals $1/2$, the total escape rate ($r_K^{noneq}$ with the factor $1/2\pi$) is half of the transition rate to the separatrix[18]. Matrix $M$ satisfies the equation (8) and the frequency factor $K(\hat{S}')$, multiplied by the frequency of excursions in the vicinity of $\hat{S}'$ in contributing to the pre-factor, satisfies the dynamics

$$\frac{dK}{dt} = -\Big[\sum_{\mu} \frac{\partial^2 H}{\partial x_\mu \partial p_\mu} + \sum_{\mu\nu} \frac{1}{2} \frac{\partial^2 S}{\partial x_\mu x_\nu} \frac{\partial^2 H}{\partial p_\mu \partial p_\nu}\Big] K \qquad (11)$$

There are two major differences in equation (7) and equation (10). (i) In equation (7), the dominant path doesn't necessary go through the saddle of the force $\hat{S}$, while in equation (10), the dominant path always go through the saddle of the force $\hat{S}$, which is a right assumption only under zero noise limit. (ii) We derived the path dependent term $e^{-S_{HJ}^{DOM}}$ from pure path integral formalism, which is an exact solution of Fokker-Plank equation. While in equation (10), the path dependent term $K(\hat{S})e^{-\int_S^{\hat{S}} \Sigma_\mu p_\mu dx_\mu}$ is derived through WKB formalism, which is just a lowest order approximated solution of Fokker-Plank equation. In zero noise approximation, the frequency factor $K(\hat{S}) \neq 1$ is from WKB approximation, which is already included in $e^{-S_{HJ}^{DOM}}$ from our path integral formalism. Therefore, it can be expected our new transition state rate will give better estimations than transition state rate of zero noise limit, especially when the fluctuation is large.

## 3 Application: Cell Fate

In this section, we will apply our new dominant path formalism and non-equilibrium transition state theory for kinetic

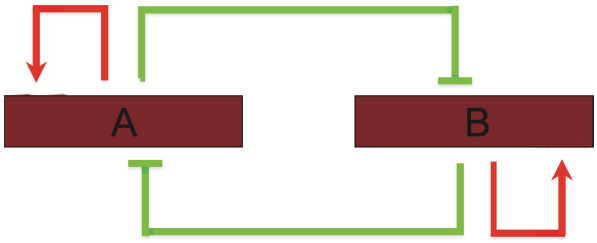



**Fig. 4** (Color online) Network diagram of canonical gene regulatory circuit of two mutually opposing proteins that positively self-regulate themselves.

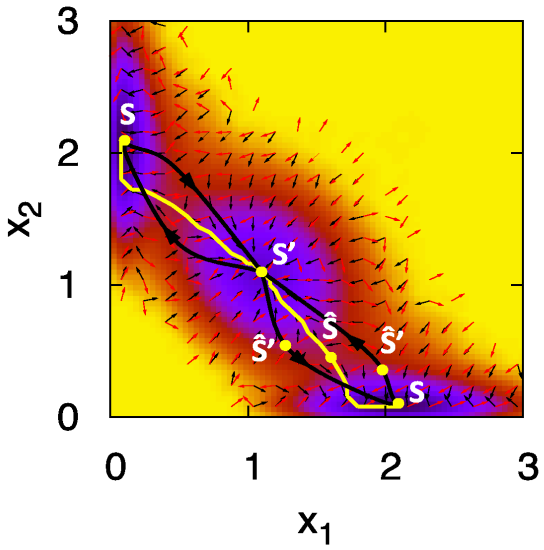



**Fig. 5** (Color online) Dominant forward ($S' \to S$ for differentiation) and backward ($S' \to S$ for reprogramming) transition paths (black lines with arrows) between differentiated states $S$ and the multipotent state $S'$ on two dimensional illustration of the underlying landscape. $\hat{S}$ is the saddle point on the landscape and $\hat{S}'$s are "the saddle points on the dominant paths". The yellow line is the gradient path along the landscape. Three blue regions represent 3 attractive basins: 2 differentiated states at $S$ and one multipotent state at $S'$. Black arrows represent the gradient force, while red arrows represent the flux.

rates to a specific example of non-equilibrium network system: a gene regulatory motif for binary cell fate decisions in stem cells. In this biological system, we found two differentiated attractors and one undifferentiated attractor. The gene regulatory circuit, as shown in Fig. 4, consists of mutual regulation of two opposing fate determining types of genes, $A$ and $B$, which can be translated into proteins $A$ and $B$ respectively. It has been shown that this module controls developmental cell fate decision (i.e. GATA1 and PU.1) in several instances of multipotent stem or progenitor cells[32,33]. The synthesis of protein $A(B)$ is controlled by the concentrations of protein $A$ and $B$. The proteins $A(B)$ can bind to the promoter of gene $A(B)$ to activate the synthesis rate of $A(B)$, which makes a self-activation feedback loop. In the meantime, protein $A(B)$ can bind to the gene $B(A)$ to repress the synthesis rate of $B(A)$, which makes a mutual repression loop.

In the adiabatic limit, the binding/unbinding processes are much faster than the synthesis/degradation, and the model can be expressed by the following chemical reactions representing the synthesis, degradation and mutual interactions of the gene products (proteins) (**supporting information**).:

$$\mathscr{O}_A \xrightarrow{g^A} A, \quad \mathscr{O}_B \xrightarrow{g^B} B, \quad A \xrightarrow{k_A} \emptyset \quad B \xrightarrow{k_B} \emptyset \tag{12}$$

in which $g^A(x_1,x_2)$ ($k_A$ ) and $g^B(x_1,x_2)$ ($k_B$) are the effective synthesis (degradation) rate of the protein $A$ and $B$ respectively. Here, $g^A(x_1,x_2)$ and $g^B(x_1,x_2)$ depend on the concentrations of the protein $A$ and $B$ ($x_1 = N_A/V, x_2 = N_B/V$) as:

$$g^A(x_1,x_2) = g_0^A + \frac{a_1 x_1^4}{S^4 + x_1^4} + \frac{b_1 S^4}{S^4 + x_2^4} \tag{13}$$

$$g^B(x_1,x_2) = g_0^B + \frac{a_2 x_2^4}{S^4 + x_2^4} + \frac{b_2 S^4}{S^4 + x_1^4} \tag{14}$$

$N_A$ and $N_B$ are molecule numbers of protein $A$ and $B$, respectively. $V$ is the cell volume. $a_1, a_2, b_1, b_2, k_A, k_B$ are positive parameters that denote the strength of the following interactions or processes: The first term represents basal level expression when there is no regulations, the second term represents a self activation (of strength $a_1, a_2$) that obeys a sigmoidal transfer function, the third term represents mutual inhibition (of strength $b_1, b_2$). Both second and third terms are determined by the thresholds and Hill coefficients characterizing the degree of cooperativity (here power 4 represents tetramer binding of regulators to the genes). Finally, the degradation of either factor is represented by the rate $k_A, k_B$. The corresponding deterministic rate equation can be given as:

$$\frac{dx_1}{dt} = g^A(x_1,x_2) - k_A x_1, \quad \frac{dx_2}{dt} = g^B(x_1,x_2) - k_B x_2 \tag{15}$$

At molecular number is finite, the intrinsic fluctuations are unavoidable. The deterministic equation above is inadequate and should be modified to the corresponding stochastic equation under fluctuations. When the number of the

molecules becomes large, the Taylor expansion to the second order leads to the Fokker-Planck equation (1) with the driving force $\vec{F} = \frac{1}{V}(g^A - k_A x_1, g^B - k_B x_2)$ and diffusion coefficients $\varepsilon_{11} = \frac{1}{V^2}(g^A + k_A x_1), \varepsilon_{22} = \frac{1}{V^2}(g^B + k_B x_2), \varepsilon_{12} = \varepsilon_{21} = 0$ where $V$ is the cell volume. However, here we do not attempt to use Fokker-Planck equation to approximate a CME. In general, no diffusion process can accurately capture both the pre-factor and the exponents of the transitions asymptotically for a CME in its limit of infinite molecule population, which is connected to Keizer's paradox. Therefore, here we only consider a diffusion process with the above driving force and the diffusion coefficients as the start of our transition state rate theory. For the comparison, we only use the Langevin dynamical simulations for diffusions instead of the Gillespie simulations for CME.

For simplicity, we consider a symmetric case for parameters of self activation, mutural repressions and degradations: $a = a_1 = a_2; b = b_1 = b_2; k = k_A = k_B$. When the parameters are set as $a = 1, b = 1, k = 1, S = 0.5, n = 4, g_0^A = g_0^B = 0.1$, it is found that there are three fixed points of the deterministic equations: two differentiated states $S_1 = (2.09481, 0.10519)$, $S_2 = (0.10519, 2.09481)$ and one undifferentiated state $S_0 = (1.1, 1.1)$. On the potential landscape, the locations of the attractors correspond to the fixed point of the averaged rate equations, as shown in Fig. 5 for $V = 25$. The two asymmetric attractors $S$ represent the differentiated states with almost mutually excluding expression of protein $A$ (i.e. GATA1) and $B$ (i.e. PU.1). On the other side, the central symmetric attractor $S'$, characterized by approximately equal expression levels of protein $A$ and protein $B$, represents the multipotent state that exhibits the characteristic balanced or promiscuous expression of the two opposing, fate-determining concentrations-a hallmark of the indeterminacy of the undecided multipotent stem cell. We also show the steady state probability flux (red arrows) on the landscape in addition to the gradient of potential landscape (black arrows). It is expected that the curl current flux component of the driving force leads to the deviation of the actual path from the one of the landscape gradient as shown in Fig. 5. Numerically, the optimal path and its weight action can be obtained by minimizing the discretized target function (**details in supporting information**).

Therefore, we can quantitatively uncover the dominant differentiation paths from the undifferentiated state $S'$ to the differentiated states at $S$, and the dominant reprogramming paths from $S$ back to $S'$, as shown in Fig. 5. It is easy to notice that the forward and backward transition paths are irreversible. In addition, neither dominant paths follow the gradient path (yellow line) on the landscape nor go through the saddle point, in contrast to what is expected from the gradient dynamics of the equilibrium systems. In a non-equilibrium system, the residual curl flux

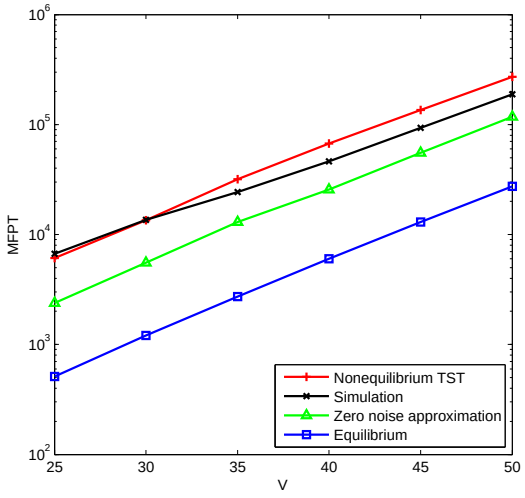



**Fig. 6** (Color online) The mean first passage time (MFPT) of the differentiation ($S' \to S$) from our theoretical predictions (Nonequilibrium TST), Langevin dynamics simulations, zero noise approximations and equilibrium transition state theory, for different cell volume $V$ (different fluctuation levels).

$$F_\mu(\vec{x})P^{SS}(\vec{x}) - \sum_\nu \frac{1}{2}\partial_\nu[\varepsilon_{\mu\nu}(\vec{x})P^{SS}(\vec{x})] = j_\mu^{SS}(\vec{x}) \qquad (16)$$

breaks the detailed balance and contributes to the weight action or the weight of the paths in a path dependent manner. The non-zero flux is contributed by the non-gradient force as the first term in equation (16). It leads to the deviation of the dominant kinetic paths from the naively expected steepest descent gradient paths and the irreversibility between the forward and the backward paths. It is worthwhile to point out that the Gaussian position dependent and non-Gaussian fluctuations can also shift the saddle and the path from passing through the original saddle. However, the forward and backward paths are reversible for these fluctuations with zero flux. For non-zero flux, the forward and backward paths are irreversible. In our example of double negative feedback network as in Fig. 4, even for the Gaussian noise with constant diffusion coefficients, the flux as defined in equation (16) is still non-zero and the system is still out-of equilibrium with irreversible paths not passing through the landscape saddle. Our double negative feedback network as in Fig. 4 provides a good example of non-equilibrium system with the non-gradient force, in which the equilibrium dominant path is inadequate and the irreversible non-equilibrium dominant paths contribute at leading order to the out-of-equilibrium kinetics.

In particular, following equation (7), since the dominant paths for the non-equilibrium systems do not pass through the

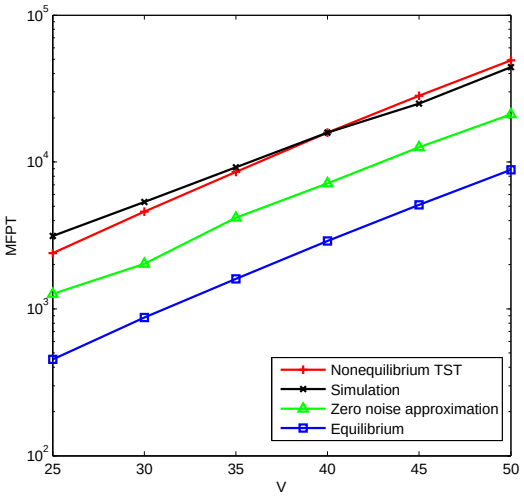



**Fig. 7** (Color online) The mean first passage time (MFPT) of the reprogramming ($S \to S'$) from our theoretical predictions (Nonequilibrium TST), Langevin dynamics simulations, zero noise approximations and equilibrium transition state theory, for different cell volume $V$ (different fluctuation levels).

transition state or the saddle point $\hat{S}$, as discussed in this study, the Kramer's formula doesn't hold any more and the kinetic transition state rate is determined by the new "saddle point" $\hat{S}'$ on the dominant path (not the saddle point on the landscape).

We then predicted accordingly the differentiation rate from the central basin of stem cell state to the side basin of differentiated cell $r_K^{noneq}$ from $S'$ to $S$, (see Fig. 6), as well as the reprogramming rate from the side basin of differentiated cell to the central basin of stem cell state $r_K^{noneq}$ from $S$ to $S'$ (see Fig. 7). The results of the kinetic transition rates $r_K^{noneq} = 1/MFPT$ from equation (7), quantifying the Nonequilibrium TST (Transition State Theory), are compared with the predictions from the Langevin dynamics simulations, the equilibrium transition state theory (Kramers' rate), as well as the zero fluctuation approximations [15,17,18]. Our theoretical predictions according to equation (7) agree with the direct stochastic simulations within the same order of magnitude for different fluctuation levels (on average, 26% larger for differentiation from $S'$ to $S$ and 3% smaller for reprogramming from $S$ to $S'$), which is better than the predictions according to the zero noise approximation (on average, 49% smaller for differentiation from $S'$ to $S$ and 55% smaller for reprogramming from $S$ to $S'$) and the prediction according to equilibrium transition state or Kramers' theory (on average, 88% smaller for differentiation from $S'$ to $S$ and 82% smaller for reprogramming from $S$ to $S'$) as shown in equation (3). As expected, zero noise approximation is worse in the small volume limit and becomes better in the large volume limit, especially for differentiation ($S'$ to $S$). Detail values can be found in tables at the end of **supporting information**. For real complex physical and biological systems in practice, the analytical rate formula presented here provides a direct and good estimation of the transition rates between different stable states.

## 4 Conclusions

In this work, we developed a new transition state or Kramers' theory and associated analytical formula for the kinetic rates from one attractor to another for general non-equilibrium dynamical systems with small but finite fluctuations. Using the weight action from the path integral, we quantify the optimal(dominant) paths. We found the optimal(dominant) paths for general non-equilibrium systems do not necessarily go through the saddle points. Importantly, we found that if we replace the saddle point by the "the saddle point on the optimal(dominant) path", the complete expression of the kinetic rate can be approximated from matching asymptotic expansions.

As a result, our new transition state theory in terms of the analytical rate formula for non-equilibrium stochastic dynamical systems is determined by the difference in weight action from the basin of attraction to the "saddle point" of the dominant kinetic paths between the two basins of the attractions in the exponential on the one hand, and by the fluctuations around the basin of attractions and the "saddle" point of the dominant kinetic path between the two basins of the attractions in the pre-factor on the other hand.

As an example, we applied our path integral formalism to a gene regulatory motif for binary cell fate decisions in stem cells. We found that the optimal paths often deviated from the gradient paths and irreversible due to the presence of the curl flux in addition to gradient force in non-equilibrium systems. Furthermore the dominant kinetic paths do not go through the saddle or transition state point on the landscape. We calculated the dynamical time scale of transition for the differentiation as an example. Our new transition state analytical rate formula is in good agreements with the stochastic simulations. Our new transition state theory and associate analytical formula for kinetic rates (or Kramers' rates) and kinetic path method are general and can be applied to other non-equilibrium biological and physical systems.

## Acknowledgement

HF and JW thank the support from NSF grant NSF-0947767 for the support. KZ thanks the National Natural Science Foundation of China (Grant no. 21190040,11174105, and 91227114) for their support. We also thank Dr. Li Xu for the help on our figures.

## References


1 Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, Pan ZJ, Schilstra MJ, Clarke PJC, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H (2002) A Genomic Regulatory Network for Development. *Science* 295:1669-1678.

2 Huang CY, Ferrell JE, Jr. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA* 93:10078-10083.

3 Wang J, Xu L, Wang EK (2008) Potential Landscape and Flux Framework of Non-Equilibrium Networks: Robustness, Dissipation and Coherence of Biochemical Oscillations. *Proc. Natl. Acad. Sci.* 105: 12271-12276.

4 Nitzan A, Ross J (1973) Oscillations, multiple steady-states, and instabilities in illuminated systems. *J. Chem. Phys.* 59:241-250.

5 Edman L, Rigler R (2000) Memory landscapes of singleenzyme molecules. *Proc. Natl. Acad. Sci.* 97:8266-8271.

6 Qian H, Elson EL (2002) Single-molecule enzymology: Stochastic Michaelis-Menten kinetics. *Biophys. Chem.* 101:565-576.

7 Qian H. (2007) Phosphorylation energy hypothesis: Open chemical systems and their biological functions. *Annu. Rev. Phys. Chem.* 58:113-142 (2007).

8 Qian H., Xie, X.S. (2006) Generalized Haldane equation and fluctuation theorem in the steady state cycle kinetics of single enzymes. *Physical Review E* 74:010902(R).

9 Min W., Jiang L., Yu, J., Kou S.C., Qian H. and Xie, X.S. (2005) Nonequilibrium steady state of a nanometric biochemical system: Determining the thermodynamic driving force from single enzyme turnover time traces. *Nano Letters* 5:2373-2378.

10 English B.P., Min W., van Oijen A.M., Lee K.T., Luo G., Sun H., Cherayil B.J., Kou S.C., Xie, X.S. (2006) Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat. Chem. Bio.* 2:87.

11 Xie X. S. (2013) Enzyme Kinetics, Past and Present. *Science* 342:1457-1459.

12 Eyring, H. (1935) The Activated Complex in Chemical Reactions. *J. Chem. Phys.* 3: 107115.

13 Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica A* 7:284-304.

14 Wang J, Zhang K, Wang EK (2010) Kinetic paths, time scale, and underlying landscapes: A path integral framework to study global natures of non-equilibrium systems and networks. *J. Chem. Phys.* 133:125103.

15 Freidlin MI, Wentzell AD (1984) Random Perturbations of Dynamical Systems. Springer-Verlag, New York/Berlin.

16 Onsager L, Machlup S (1953). "Fluctuations and Irreversible Processes". *Physical Review* 91: 15051512.

17 Schuss Z (2010) Theory and Applications of Stochastic Processes: An Analytical Approach. Springer, New York.

18 Maier RS, Stein DL, (1997) Limiting exit location distributions in the stochastic exit problem. *SIAM J. Appl. Math.* 57:752-790.

19 Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. USA* 102:18926-18931.

20 Aurell E, Sneppen K (2002) Epigenetics as a First Exit Problem. *Phys. Rev. Lett.* 88:048101.

21 Feng H, Han B, Wang J (2010) Dominant Kinetic Paths of Complex Systems: Gene Networks. *J. Phys. Chem. Lett.* 1:1836-1840.

22 Hunt KLC, Ross J (1981) Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic Lagrangian. *J. Chem. Phys.* 75:976-984.

23 Olender R, Elber R (1996) Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *J. Chem. Phys.* 105:9299-9315.

24 Wang J, Onuchic J, Wolynes PG (1996) Statistics and Kinetic Pathways of Protein Folding on Complex Energy Landscapes. *Phys. Rev. Lett.* 76:4861.

25 Wang J, Zhang K, Lu HY, Wang EK (2005) Quantifying Kinetic Paths of Protein Folding. *Biophys. J.* 89:1612-1620.

26 Faccioli P, Sega M, Pederiva F, Orland H (2006) Dominant Pathways in Protein Folding. *Phys. Rev. Lett.* 97:108101.

27 Caroli B, Caroli C, Roulet B (1981) Diffusion in a Bistable Potential: The Functional Integral Approach. *J. Stat. Phys.* 26:83-111.

28 Kalmar T, *et al.* (2009) Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *PLoS Biol.* 7:e1000149.

29 Wang J, Xu L, Wang EK, Huang S (2010) The Potential Landscape of Genetic Circuits Imposes the Arrow of Time in Stem Cell Differentiation. *Biophys. J.* 99:29-39.

30 Wang J, Zhang K, Xu L, Wang EK (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. USA* 108:8257-8262.

31 Graf T, Enver T (2009) Forcing cells to change lineages. *Nature* 462:587-594.

32 Zhou JX, Huang S (2010) Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet* 27:55-62.

33 Huang S, Guo YP, May G, Enver T (2007) Bifurcation dynamics of cell fate decision in bipotent progenitor cells. *Dev. Biol.* 305:695-713.

34 Feng H, Wang J (2012) A new mechanism of development and differentiation through slow binding/unbinding of regulatory proteins to the genes. *Scientific Reports* 2:550.

35 Danson M, Eisenthal R (2002) Enzyme assays: a practical approach. Oxford University Press, Oxford.

36 Gardiner CW (1983) Handbook of Stochastic Mechods. (Springer, New York).

37 Gillespie DT (2000) The chemical Langevin equation. *J. Chem. Phys.* 113:297-306.

38 Bialek W (2001) Stability and noise in biochemical switches. (MIT Press, Cambridge).

39 Hänggi P, Talkner P, Borkovec M (1990) Reaction Rate Theory: Fifty Years After Kramers. *Rev. Mod. Phys.* 62: 251.

40 Landauer R, Swanson J (1961) Frequency Factors in the Thermally Activated Process. *Phys. Rev.* 121: 1668.

41 Landau LD, Lifshitz EM (1996) Mechanics (3nd). (Pergamon Press, Oxford).