Quasi-potential landscape in complex multi-stable systems

**Joseph Xu Zhou[1,2,5,*], M. D. S. Aliyu[3], Erik Aurell[4,5]
and Sui Huang[1,2,5,*]**

[1]*Institute for Systems Biology, Seattle, WA, USA*
[2]*Institute for Biocomplexity and Informatics, and* [3]*Mathematics and Statistics Department,
University of Calgary, Calgary, Alberta, Canada*
[4]*Department of Computational Biology, KTH, Stockholm, Sweden*
[5]*Kavli Institute for Theoretical Physics China, CAS, Beijing, People's Republic of China*

The developmental dynamics of multicellular organisms is a process that takes place in a multi-stable system in which each attractor state represents a cell type, and attractor transitions correspond to cell differentiation paths. This new understanding has revived the idea of a quasi-potential landscape, first proposed by Waddington as a metaphor. To describe development, one is interested in the 'relative stabilities' of N attractors ($N > 2$). Existing theories of state transition between local minima on some potential landscape deal with the exit part in the transition between two attractors in pair-attractor systems but do not offer the notion of a global potential function that relates more than two attractors to each other. Several *ad hoc* methods have been used in systems biology to compute a landscape in non-gradient systems, such as gene regulatory networks. Here we present an overview of currently available methods, discuss their limitations and propose a new decomposition of vector fields that permits the computation of a quasi-potential function that is equivalent to the Freidlin–Wentzell potential but is not limited to two attractors. Several examples of decomposition are given, and the significance of such a quasi-potential function is discussed.

1. INTRODUCTION: THE NEED FOR A QUASI-POTENTIAL FUNCTION

For high-dimensional, nonlinear dynamical systems that exhibit a large number of stable steady states (attractors) far from thermodynamic equilibrium, a characteristic system behaviour of interest is manifest at the scale of the transitions between these attractors across large distances in phase space. Such systems are epitomized by gene regulatory networks of N genes, where attractor states represent the cell types in the metazoan body. Accordingly, transitions between attractor states correspond to cell phenotype changes during normal development, homeostasis and disease [1,2]. But standard analysis of dynamical systems focuses on the existence and local properties of a given steady state (linear stability). Thus, it fails to capture the essence of complex systems that operate at time scales where the transitions between attractors constitute the characteristic behaviour.

Specifically, in a system with M attractor states $\mathbf{x}_1^*, \mathbf{x}_2^*, ..., \mathbf{x}_M^*$, one would like to obtain a sense of the 'relative depth' [3] or, more precisely, the ordering of the metastable attractor states through some energy-like

function $U_1, U_2, U_3, \ldots, U_M$. Such a function $U(\mathbf{x})$ for any point $\mathbf{x}$ in the state space of the system would inform us about the probability and direction of transitions between attractor states in a noisy or perturbed system—for instance, in understanding spontaneous cell fate choice (from among a set of alternative nearby attractors for a given initial state) or the feasibility of artificial reprogramming between particular cell types. Such ordering would also explain hierarchies and the arrow of time of ontogenesis.

Let us consider a deterministic system (network) of N variables x_i (e.g. the activity of interacting genes) whose values describe the cell state $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t))^{\mathrm{T}}$ (all notations are given in table 1) and whose dynamics results from how each gene influences the activity of other genes (as invariantly predetermined by the gene regulatory network that is hard-coded in the genome). Such dynamics is described by the first-order ordinary differential equations (ODEs), which in general are nonlinear:

$$\left.\begin{array}{l} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = \boldsymbol{F}_1(x_1, x_2, ..., x_N) \\[2mm] \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = \boldsymbol{F}_2(x_1, x_2, ..., x_N) \\[2mm] \vdots \end{array}\right\}, \qquad (1.1)$$

*Authors for correspondence (joseph.zhou@systemsbiology.org; sui.huang@systemsbiology.org).

Table 1. Notation.

x, x	variables of a dynamic system $\mathbf{x} = (x_1, x_2,\ldots)$
x*	stable steady states
F(x)	driving forces of a dynamical system
$\boldsymbol{U}(\mathbf{x})$	quasi-potential function
U^L	Lyapunov function
U^{int}	proposed quasi-potential function based on integration
U^{prob}	proposed quasi-potential function based on the probability density function of steady states
U^{norm}	proposed quasi-potential function based on the normal decomposition
U^{ST}	proposed quasi-potential function based on the symmetric and anti-symmetric decomposition
$V(\boldsymbol{x})$	quasi-potential based on the least action function in the Freidlin–Wentzell theory

or in vector form:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}).$$

Such a system represents an *influence network* in which the variables (network nodes) influence each other's values according to the network rules defined by the vector $\mathbf{F}(\boldsymbol{x})$. Thus, $\mathbf{F}(\mathbf{x})$ represents the 'forces' acting to change the system state $\boldsymbol{x}(x_1, x_2, \ldots, x_N)$ in this inertia-free system. In one-dimensional systems, one always obtains a potential function $U^{\mathrm{int}}(\mathbf{x})$ by integration, which is a measure of the metastability of two (attractor) states x_A and x_B in a system:

$$U_{AB}^{\mathrm{int}}(x) = U^{\mathrm{int}}(x_B) - U^{\mathrm{int}}(x_A) = \int_{\boldsymbol{x}_A}^{x_B} \boldsymbol{F}(x)\mathrm{d}x. \quad (1.2)$$

Equivalently, the driving force is proportional to the gradient of this potential function:

$$\mathbf{F}(\mathrm{x}) \sim -\frac{\mathrm{d}\,U^{\mathrm{int}}}{\mathrm{d}x}. \quad (1.3)$$

In the system at thermodynamic equilibrium where the steady-state probability density function over $\mathbf{x}$ (giving the presence of thermal fluctuations) satisfies the Boltzmann distribution, U^{int} is the potential energy and $U_{(x)}^{\mathrm{int}} U_{AB}^{\mathrm{int}}$ is the path-independent potential difference between the two states $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. It is related to the transition probability via the Arrhenius equation,

$$P_{\boldsymbol{x}_A \to \boldsymbol{x}_B} = \mathrm{e}^{-U_{AB}^{\mathrm{int}}/\varepsilon^2}; \quad (1.4)$$

here, ε is the magnitude of noise [4].

Most works refer to transitions between equilibrium states of thermodynamic systems or simple cases such as Markovian one-dimensional non-equilibrium systems. Here we are interested in the rates of the transition between states in far-from-equilibrium, non-conservative and high-dimensional open systems with a driving force emanating from the internal interactions as described by equation (1.1). For such cases, even if we have a gradient system where a potential function U can be obtained by integration, i.e. where there exists a function $U^{\mathrm{int}}(\boldsymbol{x})$ with the following properties:

$$\frac{\partial\,U^{\mathrm{int}}}{\partial x_1} = -\boldsymbol{F}_1(\boldsymbol{x}),\ldots, \frac{\partial\,U^{\mathrm{int}}}{\partial x_i} = -\boldsymbol{F}_i(\boldsymbol{x}),\ldots, \frac{\partial\,U^{\mathrm{int}}}{\partial x_n}$$
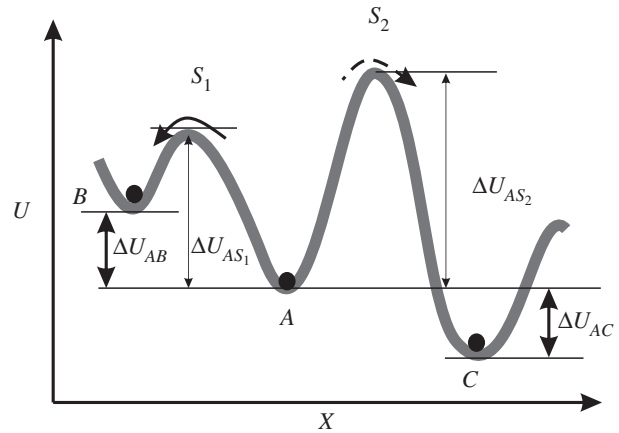$$= -\boldsymbol{F}_n(\boldsymbol{x}), \quad (1.5)$$

Figure 1. Schematic quasi-potential U and the transitions among stable steady states (attractors) in a one-dimensional multi-stable dynamical system. The transition rate $P_{A\to B}$ is not determined by ΔU_{AB} but by ΔU_{AS_1}. Similarly, the transition rate $P_{A\to C}$ is not determined by ΔU_{AC} but by ΔU_{AS_2}.

such that one can obtain, by integration, a potential function $U^{\mathrm{int}}(\boldsymbol{x})$

$$U^{\mathrm{int}}(\boldsymbol{x}) = -\int \boldsymbol{F}_1(\boldsymbol{x})\mathrm{d}x_1 + \cdots + \boldsymbol{F}_i(\boldsymbol{x})\mathrm{d}x_i + \cdots$$
$$+ \boldsymbol{F}_N(\boldsymbol{x})\mathrm{d}x_n; \quad (1.6)$$

even then the transition rate for $\boldsymbol{x}_A \to \boldsymbol{x}_B$ is not path-independent and is determined by the potentials of attractor states and saddle points between them. The transition rate $P_{\boldsymbol{x}_A \to \boldsymbol{x}_B}$ is related to $U_{AB}^{\mathrm{int}}(\boldsymbol{x}) = U^{\mathrm{int}}(\boldsymbol{x}_S) - U^{\mathrm{int}}(\boldsymbol{x}_A)$ and the transition follows the *least action path* (LAP) [5] as discussed below. Here x_{s} is the saddle point between two attractors $\mathbf{x}_A$ and $\mathbf{x}_B$, as shown in figure 1.

Yet, a potential function in non-gradient systems that is loosely referred to as 'quasi-potential' (indicated here by $\tilde{U}$ when its computation is not further specified) is not without significance if properly defined such that it can serve as a quantity for the metastability ordering of (metastable) attractor states of the system in equation (1.1). Specifically, for $\tilde{U}$ to be of meaning for our purposes in characterizing cell state transitions in development, we require that $\tilde{U}$

(i) satisfies $\mathrm{d}\tilde{U}/\mathrm{d}t < 0$ for $\boldsymbol{x} \neq \boldsymbol{x}^*$ and $\mathrm{d}\tilde{U}/\mathrm{d}t = 0$ for $\boldsymbol{x} = \boldsymbol{x}^*$, where $\boldsymbol{x}^*$ are the stable steady

states, which can be either fixed points or a limit cycle, according to the stability theory of Lyapunov [6];

(ii) is related to the Freidlin–Wentzell action along the LAP between two states and hence permits the computation of the transition rate between them [5];

(iii) we also propose that the following transitivity condition is satisfied: if $P_{a \to b} > P_{b \to a}$ and $P_{b \to c} > P_{c \to b}$ then $P_{a \to c} > P_{c \to a}$ for any three states x_a, x_b, x_c. Here $P_{i \to j}$ is the transition probability from a state x_i to another state $\boldsymbol{x}_j$.

Note that (i) is the stability property of an attractor state in a non-local sense; (ii) refers to the relationship of any two points; and only (iii) pertains to multiple attractor states. This last requirement (iii) is only conjectured to be satisfied in a subset of systems and will be elaborated in §6. This last condition must be satisfied for $\tilde{U}$ to permit global ordering of metastability as described above.

2. THE PHYSICS PROBLEM: THE QUASI-POTENTIAL FUNCTION

Because high-dimensional, non-equilibrium systems generally are not gradient systems (i.e. equation (1.5) is not satisfied),

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathbf{F}(x) \neq -\nabla \tilde{U}. \qquad (2.1)$$

By contrast, one can enforce a partial notion of a quasi-potential $\tilde{U}$, if we write the driving force as a sum of two terms,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathbf{F}(x) = -\nabla \tilde{U} + \mathbf{F}_\mathrm{r}. \qquad (2.2)$$

Where $\mathbf{F}_\mathrm{r}$ is the 'remainder' beyond the component of the driving force that follows the potential gradient. Thus, we decompose the non-gradient vector field $\mathbf{F}(\mathbf{x})$, which is finite and smooth (at least twice differentiable), into two components: one that is the gradient of some 'potential' $\tilde{U}$ and a second that represents the remainder of driving forces of the dynamic system. The question then is: what is the physical meaning of these terms? Given that there are infinite ways of decomposing a vector field into a sum of two fields, uniqueness of decomposition is imparted by the introduction of constraints, which in turn embody the physical significance. Our objective here is to find decomposition such that the quasi-potential $\tilde{U}$ presents exactly the transition barriers associated with the state transitions among steady states and the 'remainder' of the driving force $\mathbf{F}_\mathrm{r}$ will not contribute to the efforts needed for the transition.

Little work has been done to the construction of quasi-potentials given a system $\mathrm{d}x/\mathrm{d}t = \mathbf{F}(\mathbf{x})$ to determine the metastability ordering as defined above. Prigogine [7] proposed a measure of global stability in nonlinear dynamic systems far from equilibrium to compare transition barriers between any pairs of attractors

in a multi-stable system following a larger perturbation. Efimov [8] studied the mathematical properties of Lyapunov functions of multi-stable nonlinear systems but did not give the general method of how to construct them. Bhattacharya *et al.* [9] proposed a simple method to visualize a landscape in lieu of a vector field. But it represents neither Lyapunov stability nor transition rates among different attractors [9]. Ao and co-workers [10–14] suggested a general method to construct the quasi-potentials using symmetric and anti-symmetric decomposition. The method of Ao can be shown to be mathematically equivalent to our approach under certain conditions but it is not convenient for computing the transition rate between attractors in general. Xing [15] generalized Ao's work to prove that this approach is a mapping of a non-equilibrium dynamical system onto a Hamiltonian one, but misrepresented Ao's method as a generalized Helmholtz decomposition (see §3). Using stochastic systems where the dynamics manifests as the temporal evolution of the probability density function and is described by a Fokker–Planck equation, Wang *et al.* [16] championed the idea of a global potential function for gene networks using a decomposition that is similar but, again, not identical to the Helmholtz decomposition (see §3). Berglund [17] reviews Kramers' law and the computation of the transition rate in multi-stable dynamic system. He gives a rigorous mathematical discussion of the limits of applying Kramers' law in a non-gradient system [17].

Here we discuss several *ad hoc* methods used to construct or visualize a quasi-potential landscape and analyse whether they satisfy our first two conditions: (i) global metastability and (ii) agreement with the Freidlin–Wentzell theory for calculating the transition rates. Then we present the 'normal decomposition' as a systematic method that yields a quasi-potential that meets those conditions. Mathematical derivation and several examples are given to demonstrate the validity and the utility of our new approach.

3. COMPARISON OF VECTOR FIELD DECOMPOSITION

We first review several commonly used methods of vector field decompositions and the physical meaning of the components.

3.1. The Helmholtz decomposition

The *Hodge decomposition theorem* states that any sufficiently smooth (twice continuously differentiable), rapidly decaying vector field in n-dimensional space can be uniquely decomposed into the sum of a curl-free potential field $\mathrm{d}x/\mathrm{d}t = \mathbf{F}(\mathbf{x})\, U^\mathrm{H}(x)$ and a divergence-free curl term [18]. Helmholtz's theorem is a special case of Hodge decomposition theorem in three dimensions, which can be written in the following equations:

$$\left.\begin{array}{l} \mathbf{F}(\mathbf{x}) = -\nabla U^\mathrm{H}(\mathbf{x}) + \mathbf{F}_\mathrm{c}(\mathbf{x}), \\ \mathbf{F}_\mathrm{c}(\mathbf{x}) = \nabla \times \mathrm{A}(\mathbf{x}) \\ \text{and} \quad \nabla \cdot \mathbf{F}_\mathrm{c}(\mathbf{x}) = 0. \end{array}\right\} \qquad (3.1)$$
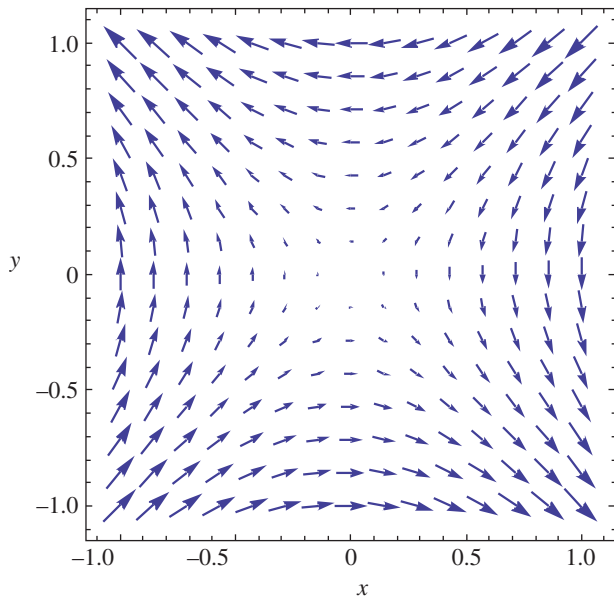
Figure 2. Curl is not necessarily a driving force for a limit circle. The vector field of a divergence-free dynamical system has open trajectories. The governing equations of the dynamical system are as follows: $dx/dt = -y$ and $dy/dt = -x$. (Online version in colour.)

The potential $U^{\mathrm{H}}(\boldsymbol{x})$ can be found by solving the following Poisson equations:

$$\nabla \cdot \boldsymbol{F}(x) = -\nabla \cdot \nabla U^{\mathrm{H}}(\mathbf{x}) + \nabla \cdot (\nabla \times A(\mathbf{x})). \quad (3.2)$$

Because the curl term is divergence-free, i.e. $(\nabla \times \boldsymbol{A}(\boldsymbol{x})) = \mathbf{0}$, we have a Poisson equation,

$$\nabla^2 U^{\mathrm{H}}(\mathbf{x}) = -\nabla \cdot \boldsymbol{F}(x). \quad (3.3)$$

Without boundary conditions, the decomposition is not unique because any specific solution that satisfies equation (3.3) plus any harmonic function will also be qualified. But the decomposition is unique with the boundary conditions that $\mathbf{F}(\mathbf{x}) \rightarrow \mathbf{0}$ when $\boldsymbol{x}$ approaches infinity. Therefore, the potential $U^{\mathrm{H}}(\boldsymbol{x})$ can be directly 'read off' from the dynamic equations (1.1) without a time-stepping process by solving the static Poisson equations with known right-hand terms and boundary conditions.

The meaning of this decomposition is that any dynamical field can be divided into two parts: a conservative irrotational field whose gradient connects the attractors or repellors (analogous to sinks and sources in the field theory) and a solenoidal curl field that is not influenced by sinks or sources. However, given that the divergence-free curl field has been used to explain the driving force of the gradient-free limit cycle [19], it is noteworthy that the curl field does not necessarily consist of closed trajectories. It can be made up of open trajectories with varying direction, as shown in figure 2. Nevertheless, this decomposition permits the study of the factors that influence the period of limit cycles (the curl term) and their stability (the gradient term) separately.

We can now determine whether the potential function obtained from the Helmholtz decomposition, U^{H}, satisfies the Lyapunov criterion for stability,

our requirement (i), i.e. in particular, if $dU^{\mathrm{H}}/dt < 0$ for $\boldsymbol{x} \neq \boldsymbol{x}^*$. For a two-dimensional system taking the time derivative of U^{H} from equation (3.1), we obtain

$$\frac{dU^{\mathrm{H}}}{dt} = -\left(\frac{\partial U^{\mathrm{H}}}{\partial x}\right)^2 - \left(\frac{\partial U^{\mathrm{H}}}{\partial y}\right)^2 + (\nabla U^{\mathrm{H}}, \mathbf{F}_{\mathrm{c}}). \quad (3.4)$$

Because we only know that $\nabla \cdot \mathbf{F}_{\mathbf{c}}(\boldsymbol{x}) = 0$, $(\nabla U^{\mathrm{H}}, \mathbf{F}_{\mathrm{c}})$ can be either positive or negative. Thus, dU^{H}/dt is not necessarily non-positive, and hence Lyapunov stability is not guaranteed in this decomposition. Therefore, U^{H} does not satisfy our criterion (i) and it is not a measure of global metastability.

3.2. Decomposition based on the flux of the probability, $\mathbf{U} \sim -ln\mathbf{P}$

An often taken, *ad hoc* approach to obtain a quasi-potential function $\tilde{U}$ and to visualize it as a landscape [16,20,21] stems from the intuitive notion that, in a probabilistic form of the system, the more stable states (lower potential $\tilde{U}$) are more probable in the state space. Thus, the steady-state probability for each state x is then linked to a potential U^{prob}, according to the Boltzmann law

$$U^{\mathrm{prob}}(x, t) \sim -\ln(P(\boldsymbol{x}, t)). \quad (3.5)$$

Again, because in general $\mathbf{F}(\boldsymbol{x}) \neq \nabla U^{\mathrm{prob}}$, the driving force $\mathbf{F}(\boldsymbol{x})$ requires an additional component besides the gradient term. With the constraints of equation (3.5), i.e. a gradient force emanating from a probability flux as one component, Wang *et al.* [22] derived the following decomposition.

The time evolution of the probability density function $P(\boldsymbol{x}, t)$ is governed by the Fokker–Planck equation, which can accurately describe the transition dynamics in a multi-stable dynamic system if the noise is Gaussian distribution,

$$\frac{\partial P(\boldsymbol{x}, t)}{\partial t} = -\nabla \cdot (\mathbf{F}P) + D \cdot \nabla^2 P, \quad (3.6)$$

in which the drift term is the driving force $\mathbf{F}(x)$ of our system. Note that the diffusion term D needs to be large enough for the probability density function $P(\boldsymbol{x}, t)$ to achieve a good coverage of the whole state space. The temporal change of the probability density can be written in terms of the probability flux $\mathbf{J}(\boldsymbol{x}, t)$,

$$\left. \begin{array}{l} \dfrac{\partial P(\boldsymbol{x}, t)}{\partial t} = -\nabla \cdot \mathbf{J}(x, t) \\[2mm] \text{and} \quad \mathbf{J}(\boldsymbol{x}, t) = \boldsymbol{F}\,P - D \cdot \nabla P. \end{array} \right\} \quad (3.7)$$

Because at steady states $\partial P(\boldsymbol{x}, t)/\partial t = 0$ the divergence of the steady-state probability flux $\boldsymbol{J}_{\mathrm{ss}}$ vanishes, i.e. $\nabla \cdot \boldsymbol{J}_{\mathrm{ss}}(\boldsymbol{x}, t) = 0$, we obtain

$$\begin{aligned} \mathbf{F} &= \frac{\boldsymbol{J}_{\mathrm{ss}}}{P_{\mathrm{ss}}} + D \cdot \frac{\nabla P_{\mathrm{ss}}}{P_{\mathrm{ss}}} \\[2mm] &= -\nabla - \ln(P_{\mathrm{ss}}) \cdot D + \frac{J_{\mathrm{ss}}}{P_{\mathrm{ss}}} \\[2mm] &= -\nabla U^{\mathrm{prob}} + \mathbf{F}_{\mathrm{c}}. \end{aligned} \quad (3.8)$$

In this vector field decomposition based on steady-state probability density, the driving force $\mathbf{F}(\boldsymbol{x})$ has

been decomposed into two terms,

$$
\left.\begin{aligned}
-\nabla U^{\mathrm{prob}} &= -\nabla(-\ln(P_{\mathrm{ss}})) \cdot D \\
\text{and} \quad \mathbf{F}_c &= \frac{J_{\mathrm{ss}}}{P_{\mathrm{ss}}}.
\end{aligned}\right\} \tag{3.9}
$$

Thus, the first term consists indeed of a gradient of a potential U^{prop}. The second term is the flux $J_{\mathrm{ss}}/P_{\mathrm{ss}}$ and has a curl nature. However, it is not a curl in the same sense as in the Helmholtz decomposition because generally it is not divergence-free: $\nabla \cdot (J_{\mathrm{ss}}/P_{\mathrm{ss}}) \neq 0$. Only J_{ss} itself is divergence-free. Thus this decomposition is not a Helmholtz decomposition as claimed elsewhere [21].

To test whether U^{prob} of this decomposition meets both our requirement (i) and our requirement (ii), we insert $P = (\mathrm{e}^{-U^{\mathrm{prob}}/D})$ and $\mathbf{F} = -\nabla U^{\mathrm{prob}} + \mathbf{F}_c$ into the Fokker–Planck equation,

$$
\left.\begin{aligned}
\frac{\partial P}{\partial t} &= -\nabla \cdot (\mathbf{F}P) + D \cdot \nabla^2 P \\
\text{and} \\
\frac{\partial}{\partial t}(\mathrm{e}^{-U^{\mathrm{prob}}/D}) &= -\nabla((-\nabla U^{\mathrm{prob}} + \mathbf{F}_c) \cdot \mathrm{e}^{-U^{\mathrm{prob}}/D}) \\
&\quad + D \cdot \nabla^2 \mathrm{e}^{-U^{\mathrm{prob}}/D},
\end{aligned}\right\} \tag{3.10}
$$

and obtain

$$
(-\nabla U^{\mathrm{prob}}, \mathbf{F}_c) = D\,\nabla \cdot \mathbf{F}_c. \tag{3.11}
$$

That is, the two vector fields are not perpendicular unless $\mathbf{F}_c = J_{\mathrm{ss}}/P_{\mathrm{ss}}$ is divergence-free—which is not necessarily so, as argued above. In other words, this decomposition, while motivated by $U^{\mathrm{prob}} \sim -\ln(P)$, is actually a field decomposition $\mathbf{F} = -\nabla U + \mathbf{F}_c$ with the constraint $(-\nabla U^{\mathrm{prob}}, \mathbf{F}_c) = D\nabla \cdot \mathbf{F}_c$.

We can now again determine whether the potential function obtained, $U^{\mathrm{prob}} \sim -\ln(P_{\mathrm{ss}})$, satisfies the Lyapunov criterion for stability, our requirement (i), i.e. in particular, if $\mathrm{d}U^{\mathrm{prob}}/\mathrm{d}t < 0$ for $\boldsymbol{x} \neq \boldsymbol{x}^*$. For a two-dimensional system taking the time derivative of U^{prob}, we obtain

$$
\begin{aligned}
\frac{\mathrm{d}U^{\mathrm{prob}}}{\mathrm{d}t} &= -\left(\frac{\partial U^{\mathrm{prob}}}{\partial x}\right)^2 - \left(\frac{\partial U^{\mathrm{prob}}}{\partial y}\right)^2 + (\nabla U^{\mathrm{prob}}, \mathbf{F}_c) \\
&= -\left(\frac{\partial U^{\mathrm{prob}}}{\partial x}\right)^2 - \left(\frac{\partial U^{\mathrm{prob}}}{\partial y}\right)^2 + (D\nabla \cdot \mathbf{F}_c).
\end{aligned} \tag{3.12}
$$

Thus, $\mathrm{d}U^{\mathrm{prob}}/\mathrm{d}t$ is not necessarily non-positive, and hence Lyapunov stability is not guaranteed in this decomposition. Therefore, generally speaking, $U^{\mathrm{prob}} \sim -\ln(P_{\mathrm{ss}})$ cannot be considered a measure of global metastability. One argument to save $U^{\mathrm{prob}} \sim -\ln(P)$ is that equation (3.11) will hold if the noise term D goes to zero in the limit, i.e. if $D \to 0$ then $D\nabla \cdot \mathbf{F}_c \to 0$. It implies that the gradient component and the remainder $\mathbf{F}_c$ would be perpendicular to each other and that the Lyapunov criterion is also satisfied. However, the noise term D cannot be too small. If D indeed goes to zero, state transitions between attractors will never happen and P_{ss} will not be a global probability

distribution but only a probability distribution in local attractors as a result of the balance between noise and driving forces.

We next examine the meaning for the case when the constraint is such that the gradient force and the remainder force are perpendicular to each other because this would guarantee satisfaction of the Lyapunov stability criterion.

3.3. The normal decomposition

From the discussion above, the constraint that the gradient term is perpendicular to the 'remainder' force $\mathbf{F}_\perp$ is of interest,

$$
\mathbf{F} = -\nabla U^{\mathrm{norm}} + \mathbf{F}_\perp. \tag{3.13}
$$

Taking the same approach as in equation (3.12), for a two-dimensional system, we obtain the time derivative for U^{norm} along the trajectory. Supposing that we have a Lyapunov function U for a two-dimensional system, its time derivative along any trajectory driven by the dynamic system in equation (1.1) is

$$
\begin{aligned}
\frac{\mathrm{d}U^{\mathrm{norm}}}{\mathrm{d}t} &= \frac{\partial U^{\mathrm{norm}}}{\partial x}\dot{x} + \frac{\partial U^{\mathrm{norm}}}{\partial y}\dot{y} \\
&= \frac{\partial U^{\mathrm{norm}}}{\partial x(-\partial U^{\mathrm{norm}}/\partial x + \boldsymbol{F}_\perp^x)} + \frac{\partial U^{\mathrm{norm}}}{\partial y(-\partial U^{\mathrm{norm}}/\partial y + \boldsymbol{F}_\perp^y)} \\
&= -\left(\frac{\partial U^{\mathrm{norm}}}{\partial x}\right)^2 - \left(\frac{\partial U^{\mathrm{norm}}}{\partial y}\right)^2 + (\nabla U^{\mathrm{norm}}, \mathbf{F}_\perp). \tag{3.14}
\end{aligned}
$$

If the gradient of U^{norm} is normal to the remainder force $\mathbf{F}_\perp$, i.e.

$$
(\nabla U^{\mathrm{norm}}, \mathbf{F}_\perp) = 0, \tag{3.15}
$$

then the function U^{norm} will decrease monotonically with time during the process before reaching the steady states, thus satisfying *Lyapunov's* condition for global metastability,

$$
\frac{\mathrm{d}U^{\mathrm{norm}}}{\mathrm{d}t} = -\left(\frac{\partial U^{\mathrm{norm}}}{\partial x}\right)^2 - \left(\frac{\partial U^{\mathrm{norm}}}{\partial y}\right)^2 < 0. \tag{3.16}
$$

We can therefore see that the condition for the normal decomposition has an obvious physical meaning, in that, for $(\nabla U^{\mathrm{norm}}, \mathbf{F}_\perp) = 0$, U^{norm} corresponds to a Lyapunov function U^{norm} of dynamical systems and can represent the global metastability. Thus, we seek a decomposition of a sufficiently smooth vector field into a conservative potential field U^{norm} and the remaining forces $\mathbf{F}_\perp$,

$$
\left.\begin{aligned}
\mathbf{F} &= -\nabla U^{\mathrm{norm}} + \mathbf{F}_\perp \\
(-\nabla U^{\mathrm{norm}}, \mathbf{F}_\perp) &= 0.
\end{aligned}\right\} \tag{3.17}
$$

The physical interpretation is that for a ball in an attractor state that is perturbed to exit with least 'energy' against the system's 'driving force' $\mathbf{F}(\boldsymbol{x})$ that keeps it in the attractor, we can decompose the field such that $\mathbf{F}_\perp$ will not contribute to this process but only a gradient field U^{norm} contributes. On the basis of Freidlin–Wentzell's large deviation theory of a stochastic process discussed below [5], U^{norm} can thus be used to compute the transition rate in a non-equilibrium dynamic system.

The normal potential U^{norm} can also be read off the systems equations without a time-stepping solution. We can calculate the potential field U^{norm} as follows:

$$(\nabla U^{\mathrm{norm}}, \mathbf{F} + \nabla U^{\mathrm{norm}}) = 0. \qquad (3.18)$$

This equation is the Hamilton–Jacobi equation (HJE), which can be written in a component format

$$\frac{\partial U^{\mathrm{norm}}}{\partial x_1} \cdot \left(\boldsymbol{F}_{x_1} + \frac{\partial U^{\mathrm{norm}}}{\partial x_1} \right) + \cdots + \frac{\partial U^{\mathrm{norm}}}{\partial x_i}$$
$$\cdot \left(\boldsymbol{F}_{x_i} + \frac{\partial U^{\mathrm{norm}}}{\partial x_i} \right) + \cdots + \frac{\partial U^{\mathrm{norm}}}{\partial x_n}$$
$$\cdot \left(\boldsymbol{F}_{x_n} + \frac{\partial U^{\mathrm{norm}}}{\partial x_n} \right) = 0. \qquad (3.19)$$

The HJE provides for any non-equilibrium dynamic systems a general mathematical framework to construct a quasi-potential function that allows us to compare the relative stability of different attractors. This extends the global stability analysis of Lyapunov's theory to multi-stable dynamic systems. The quasi-potential function U^{norm} also epitomizes the epigenetic landscape that could explain the driving force behind cell differentiation in multicellular organisms, which can be used to calculate the transition rates between different cellular attractors (more details in §4 and §6)

Note that the existence of a solution for equation (3.19) is by no means guaranteed for all dynamic systems. However, here we deal only with a subset of problems encountered in biology rather than a general dynamic system. We now propose that the dynamical systems that represent organismal development admit such a decomposition. Further research is needed to find the conditions for the existence of such a decomposition and verify them in the real biological systems.

It is far from trivial to solve a nonlinear PDE such as the HJE, which is still a topic of active research in applied mathematics. Since there usually exist no analytical solutions, numerical methods with careful convergence and stability analysis have to be applied to solve the HJE in the real biological systems. A numerical method, the Newton–Raphson method, is used in this paper, which is described in the electronic supplementary material.

3.4 The relationship between the normal decomposition and the symmetric–antisymmetric decomposition

The symmetric–antisymmetric decomposition proposed by Ao *et al.* [23] is under certain conditions mathematically equivalent to the normal decomposition approach presented above.

3.4.1. The symmetric–antisymmetric decomposition when D = I
For a dynamic system with governing equations in equation (1.1), the symmetric–antisymmetric

decomposition can be achieved as follows [14]:

$$\left. \begin{array}{l} (\boldsymbol{S} + \boldsymbol{T}) \cdot \boldsymbol{F} = -\nabla U^{ST} \\ \text{and} \qquad (\boldsymbol{S} - \boldsymbol{T})\boldsymbol{D}(\boldsymbol{S} + \boldsymbol{T}) = \boldsymbol{S}, \end{array} \right\} \qquad (3.20)$$

where U^{ST} is the quasi-potential derived from the symmetric–antisymmetric decomposition. $\boldsymbol{S}$ is a symmetric operator, $\boldsymbol{T}$ is an antisymmetric operator. Because it is usually difficult to find the analytical solutions for $\boldsymbol{S}$ and $\boldsymbol{T}$, $\boldsymbol{S}$ *can be solved as a* symmetric matrix and $\boldsymbol{T}$ as an antisymmetric matrix in finite dimensional approximation. $\boldsymbol{D}$ is a diffusion matrix. Equation (1.1) can be rewritten as the following:

$$\boldsymbol{F} = -\nabla U^{ST} + \boldsymbol{F}_{\mathrm{r}} = (S + T) \cdot \boldsymbol{F} + \boldsymbol{F}_{\mathrm{r}}. \qquad (3.21)$$

Then $\boldsymbol{F}_{\mathrm{r}}$ can be written as

$$\boldsymbol{F}_{\mathrm{r}} = (\boldsymbol{I} - \boldsymbol{S} - \boldsymbol{T}) \cdot \boldsymbol{F}. \qquad (3.22)$$

$\boldsymbol{I}$ is an identity matrix.

To follow our perpendicular decomposition, the inner product of two terms is

$$\begin{aligned} (\nabla U^{ST}, \boldsymbol{F}_{\mathrm{r}}) &= ((\boldsymbol{S} + \boldsymbol{T}) \cdot \boldsymbol{F}, (\boldsymbol{I} - \boldsymbol{S} - \boldsymbol{T}) \cdot \boldsymbol{F}) \\ &= \boldsymbol{F}^{\mathrm{T}} (\boldsymbol{S} + \boldsymbol{T})^{\mathrm{T}} (\boldsymbol{I} - (\boldsymbol{S} + \boldsymbol{T})) \boldsymbol{F} \qquad (3.23) \\ &= \boldsymbol{F}^{\mathrm{T}} (\boldsymbol{S} - (\boldsymbol{S} - \boldsymbol{T})(\boldsymbol{S} + \boldsymbol{T})) \boldsymbol{F}. \end{aligned}$$

If $\boldsymbol{D}$ is the identify matrix, i.e. $\boldsymbol{D} = \mathrm{I}$, then, from equation (3.20), we have

$$(\boldsymbol{S} - \boldsymbol{T})(\boldsymbol{S} + \boldsymbol{T}) = \boldsymbol{S}. \qquad (3.24)$$

And from equations (3.23) and (3.24), it is easy to show that $(\nabla U^{ST}, \boldsymbol{F}_{\mathrm{r}}) = 0$, i.e. the gradient of the quasi-potential ∇U^{ST} is perpendicular to $\boldsymbol{F}_{\mathrm{r}}$. Therefore, the symmetric–antisymmetric decomposition is mathematically identical to our normal decomposition if the diffusion matrix $\boldsymbol{D}$ is set to be an identity matrix $\boldsymbol{I}$.

3.4.2. Symmetric–antisymmetric decomposition when D ≠ I
The ensuing question is what the quasi-potential U^{ST} obtained from the symmetric–antisymmetric decomposition will be if $\boldsymbol{D} \neq \boldsymbol{I}$. We can again evaluate this using the two conditions that we proposed in §2. First, we find that U^{ST} decreases monotonically with time during the approach to the steady states, thus satisfying Lyapunov's condition for global metastability:

$$\begin{aligned} \frac{\mathrm{d}U^{ST}}{\mathrm{d}t} &= \nabla U^{ST} \boldsymbol{F} = -\boldsymbol{F}^T (\boldsymbol{S} + \boldsymbol{T}) \boldsymbol{F} \\ &= -\boldsymbol{F}^T \boldsymbol{S} \boldsymbol{F} < 0. \qquad (3.25) \end{aligned}$$

We can therefore see whether the solution for equation (3.20) exists; U^{ST} is indeed a Lyapunov function of a dynamical system that can represent global metastability.

However, if $\boldsymbol{D}$ is not an identity matrix $\boldsymbol{I}$, i.e. $(\nabla U^{ST}, \boldsymbol{F}_{\mathrm{r}}) \neq 0$, the remainder $\mathbf{F}_{\mathrm{r}}$ is not perpendicular to ∇U^{ST} and thus will contribute to the state transition. Then, on the basis of the Freidlin–Wentzell large deviation theory of a stochastic process (discussed below) [5], U^{ST} cannot be used to compute the transition rate in a non-equilibrium dynamic system.

4. FREIDLIN–WENTZELL THEORY OF LARGE DEVIATION IN MULTI-STABLE SYSTEMS AND ITS RELATIONSHIP WITH THE NORMAL DECOMPOSITION

4.1. Background

For a dynamic system with deterministic forces $\mathbf{F}(\boldsymbol{x},t)$,

$$\frac{\mathrm{d}\boldsymbol{x0}}{\mathrm{d}t} = \mathbf{F}(\boldsymbol{x0}, t). \qquad (4.1)$$

Now let us consider that the system is under a stochastic perturbation $\boldsymbol{\varepsilon}$,

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \mathbf{F}(\boldsymbol{x}, t) + \boldsymbol{\varepsilon} \cdot \boldsymbol{\xi}(t). \qquad (4.2)$$

If $\boldsymbol{\varepsilon}$ is sufficiently small, the perturbed system will converge to the original dynamical system, i.e. $\|\boldsymbol{x} - \boldsymbol{x0}\| \to 0$. However, if the perturbation is a random process with small average amplitude but with occasional large excursions, the perturbed dynamic system will behave differently. Freidlin & Wentzell [5] proposed a large deviation theory of stochastic process as a theoretical framework for the analysis of dynamical systems with multiple attractors. Supposing that a dynamical system satisfies the Langevin dynamics, the governing equations are described by the following ODEs:

$$\left. \begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n) + \xi_1(t) \\ &\vdots \\ \dot{x}_i &= f_i(x_1, \ldots, x_n) + \xi_i(t) \\ &\vdots \\ \dot{x}_1 &= f_{n1}(x_1, \ldots, x_n) + \xi_n(t). \end{aligned} \right\} \qquad (4.3)$$

Supposing that a ball is perturbed to go from state $\mathbf{x}_A^* \to \mathbf{x}_B^*$ by a stochastic process, one defines an action function V_{AB} to measure the 'energy' barrier to be overcome for this transition,

$$V_{AB} = \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\sum_{i=1}^{n}\|\mathbf{k}_1 - f_i(\mathbf{x})\|^2\right]\mathrm{d}t\right\}. \qquad (4.4)$$

Here, the action function V_{AB} is defined as a time integral of the square of the 'remainder' of the dynamic equations (deviation from deterministic trajectory) over the whole trajectory $X(t)$ from attractor $\mathbf{x_A}$ to $\mathbf{x_B}$. If a ball is only driven by the 'forces' specified in the deterministic part of the ODEs (equation 4.2), it will correspond to a 'free fall' in the 'gravity field' and the action is zero. If a ball is perturbed against the 'forces', the remainder term $\|\dot{x}_i - f_i(\boldsymbol{x})\|^2$ will not be zero. We integrate them over time along the whole trajectory and obtain the total action when the ball switches from attractor $\mathbf{x_A}$ to $\mathbf{x_B}$. On the basis of the *variation principle*, there exists a unique minimum action $V_{\mathbf{AB}}$, which is an objective measure of the difficulty for the state transition in non-equilibrium dynamic systems.

4.2. The relationship between the Freidlin–Wentzell potential and normal decomposition

Although the Freidlin–Wentzell potential V and the normal potential U^{norm} are defined in different ways, they are mathematically related. For a dynamic system, $\mathbf{F} = -\nabla U^{\mathrm{norm}} + \mathbf{F}_{\perp}$ with $(-\nabla U^{\mathrm{norm}}, \mathbf{F}_{\perp}) = 0$. We can rewrite the Wentzell potential as

$$\begin{aligned} V_{AB} &= \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\|\mathbf{X} - \mathbf{F}\|^2\right]\mathrm{d}t\right\} \\ &= \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\|\mathbf{X} + \nabla U^{\mathrm{norm}} - \mathbf{F}_{\perp}\|^2\right]\mathrm{d}t\right\} \\ &= \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\|\mathbf{X} - \mathbf{F}_{\perp} - \nabla U^{\mathrm{norm}}\|^2\right]\mathrm{d}t \right. \\ &\qquad \left. + 4\int_{tA}^{tB}\left[(\mathbf{X} - \mathbf{F}_{\perp}) - \nabla U^{\mathrm{norm}}\right]\mathrm{d}t\right\} \\ &= \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\|\mathbf{X} - \mathbf{F}_{\perp} - \nabla U^{\mathrm{norm}}\|^2\right]\mathrm{d}t \right. \\ &\qquad \left. + 4\int_{tA}^{tB}\left[\mathbf{X} - \nabla U^{\mathrm{norm}}\right]\mathrm{d}t\right\} \\ &= \frac{1}{2}\min\left\{\int_{tA}^{tB}\left[\|\mathbf{X} - \mathbf{F}_{\perp} - \nabla U^{\mathrm{norm}}\|^2\right]\mathrm{d}t \right. \\ &\qquad \left. + 4(U_B^{\mathrm{norm}} - U_A^{\mathrm{norm}})\right\} \\ &= 2(U_B^{\mathrm{norm}} - U_A^{\mathrm{norm}}). \end{aligned} \qquad (4.5)$$

During the process of exiting an attractor, the LAP of a ball follows the governing equation:

$$\mathbf{F} = \nabla U^{\mathrm{norm}} + \mathbf{F}_{\perp}. \qquad (4.6)$$

Intuitively, the LAP out of attractors will be driven by a landscape that reverses the original one, i.e. the hills are reverted to valleys, and valleys are changed to be hills. However, the reversal only holds for the gradient component because the normal remainder $\mathbf{F}_{\perp}$ still retain the same direction as before [24]. As long as we are within the same attractor, the Freidlin–Wentzell potential is exactly twice as big as the potential from the normal decomposition, $V_{AB} = 2(U_B^{\mathrm{norm}} - U_A^{\mathrm{norm}})$.

When a 'ball' transitions from one attractor to another, the Freidlin–Wentzell potential only accounts for the uphill 'energy', which is two times that of U^{norm}. Once it goes over the 'saddle' point (S_1 and S_2 in figure 1), the Wentzell potential V is zero for the remaining free-fall path. The same applies for a ball that transitions through *many* attractors in between: the Wentzell potential is equal to the sum of all uphill potentials between two points. All downhill paths contribute nothing to the Freidlin–Wentzell potential.

4.3. Wentzell potential and spontaneous transition rate

For a system that is under a large deviation perturbation: $\mathrm{d}x/\mathrm{d}t = \mathbf{F}(\mathbf{x},t) + \boldsymbol{\varepsilon}\cdot\boldsymbol{\xi}(t)$, V_{AB}, as defined in

equation (4.4), the probability of spontaneous transition $\mathbf{x}_A \rightarrow \mathbf{x}_B$, is [5]

$$P_{\boldsymbol{x}_A \rightarrow x_B} = \mathrm{e}^{-V_{AB}/\varepsilon^2}. \tag{4.7}$$

Accordingly, for the transition between $\boldsymbol{x}_B \rightarrow \boldsymbol{x}_A$ the probability is

$$P_{x_B \rightarrow x_A} = \mathrm{e}^{-V_{BA}/\varepsilon^2}. \tag{4.8}$$

If $V_{AB} > V_{BA}$ under random perturbation, a ball will stay longer in state $\mathbf{x}_B$, i.e. state $\mathbf{x}_B$ is globally more stable than state $\mathbf{x}_A$ and, thus, the directionality of the spontaneous transition points from state $\mathbf{x}_A$ to state $\mathbf{x}_B$.

4.4. Wentzell potential and least action path of spontaneous transition

Although the Wentzell potential V_{AB} can be mathematically expressed as the normal potential U^{norm} it offers additional information that cannot be directly derived from U^{norm}. In determining V_{AB}, not only a minimum value is computed, but also a LAP X_{AB} between two points is obtained.

However, even if there is a LAP, a state transition $\mathbf{x}_A \rightarrow \mathbf{x}_B$ does not necessarily have to follow exactly this path only. Because the perturbation is random, the effective state transition can happen along any path that topologically connects these two states. The majority of probable paths fall near the area of the LAP, i.e. for any path $\mathbf{X_t}(\mathbf{x}_A \rightarrow \mathbf{x}_B)$, the LAP is $\mathbf{X_{AB}}$ and

$$P(\| \mathbf{X_t} - \mathbf{X_{AB}} \| < \varepsilon) = \mathrm{e}^{-V_{AB}/\varepsilon^2}. \tag{4.9}$$

Here, $\| \ \|$ is defined as a norm, $\| X \| = \left(\sum_{i=1}^{n}(x_i)^2\right)^{1/2}$. It means that the bigger the noise, the further away the actual transition paths can deviate from the LAP $\mathbf{X_{AB}}$.

5. EXAMPLES

In this section, two examples are presented to show which decomposition can give the correct transition rates between metastable attractors in nonlinear, high-dimensional dynamical systems. Quasi-potentials $U(\boldsymbol{x})$ are computed using different decompositions: the Helmholtz decomposition (§3.1), the $U \sim -\ln P$ decomposition (§3.2) and the normal decomposition (§3.3). The transition barriers ΔU are calculated based on the quasi-potentials constructed from these decompositions. The results are compared with the Wentzell action functions that represent the correct transition rates among different attractors.

5.1. Example 1. Spiral dynamic system with a single attractor

A two-dimensional non-gradient dynamical system reveals the effects of the non-vanishing remainder component of the vector field. The governing equations are

$$\left.\begin{array}{l}\dot{x} = -x + y^2 \\ \dot{y} = -y - xy.\end{array}\right\} \tag{5.1}$$

The perpendicular decomposition of the vector field can be easily found from equations (3.18) and (3.19),

$$\left.\begin{array}{l}\dot{x} = -\dfrac{\partial U^{\mathrm{norm}}}{\partial x} + \boldsymbol{F}_{\perp}^x = -\dfrac{\partial(x^2+y^2)/2}{\partial x} + y^2 \\[3mm] \dot{y} = -\dfrac{\partial U^{\mathrm{norm}}}{\partial x} + \boldsymbol{F}_{\perp}^y = -\dfrac{\partial(x^2+y^2)/2}{\partial y} - xy.\end{array}\right\} \tag{5.2}$$

The quasi-potential functions and the remainders from the perpendicular decomposition are

$$\begin{array}{l}U^{\mathrm{norm}} = \dfrac{(x^2+y^2)}{2} \\[3mm] \text{and} \quad \left\{\begin{array}{l}\boldsymbol{F}_{\perp}^x = y^2 \\ \boldsymbol{F}_{\perp}^y = -xy.\end{array}\right.\end{array}\right\} \tag{5.3}$$

It can be easily verified that

$$\frac{\partial U^{\mathrm{norm}}}{\partial x} \cdot \boldsymbol{F}_{\perp}^x + \frac{\partial U^{\mathrm{norm}}}{\partial y} \cdot \boldsymbol{F}_{\perp}^y = x \cdot y^2 - y \cdot xy = 0. \tag{5.4}$$

The Helmholtz decomposition can be constructed in a straightforward manner,

$$\left.\begin{array}{l}\dot{x} = -\dfrac{\partial U^{\mathrm{H}}}{\partial x} + \boldsymbol{F}_{\mathrm{c}}^x = -\dfrac{\partial(x^2+y^2+xy^2)/2}{\partial x} + \dfrac{3}{2}y^2 \\[3mm] \dot{y} = -\dfrac{\partial U^{\mathrm{H}}}{\partial x} + \boldsymbol{F}_{\mathrm{c}}^y = -\dfrac{\partial(x^2+y^2+xy^2)/2}{\partial y}.\end{array}\right\} \tag{5.5}$$

We have

$$\begin{array}{l}U^{\mathrm{H}} = (x^2+y^2+xy^2)/2 \\[3mm] \text{and} \quad \left\{\begin{array}{l}\boldsymbol{F}_c^x = \dfrac{3}{2}y^2 \\ \boldsymbol{F}_c^y = 0.\end{array}\right.\end{array}\right\} \tag{5.6}$$

As mentioned before, the Helmholtz decomposition is not unique. The superposition of the function above with any harmonic function still satisfies the Helmholtz decomposition. However, this example illustrates the general characteristics of Helmholtz decomposition: the remainder and the gradient components are not necessarily perpendicular; thus, according to §4, it is not the right decomposition to give the correct transition rate.

Because U^{prob} in the $-\ln P$ decomposition (§3.2) depends on solving the steady states of the Fokker–Planck equation, which usually does not have an analytical solution, here we calculate $U \sim -\ln P$ using a numerical method (finite difference method); see details in the electronic supplementary material. Different noise levels in the Fokker–Planck equation lead to different distributions $P(\mathbf{x})$ in steady states. $U^{\mathrm{prob}} = -\ln P \times D$ were calculated with different noise levels ($D = 0.015$, 0.045, 0.09) to cover a wider range of results from the $-\ln P$ decomposition.

The quasi-potentials at nine points (relative to the origin, i.e. $U(0,0) = 0$) from four different methods were calculated based on the methods described above and shown in figure 3. As can be recognized, the quasi-potential U^{H} (shown by circles) from the Helmholtz decomposition significantly departs from the Wentzell potential, and thus is not good for calculating the transition probability. The quasi-potential from the $-\ln P$ decomposition U^{Prob} has two problems. First, if the noise level D is small ($D =$
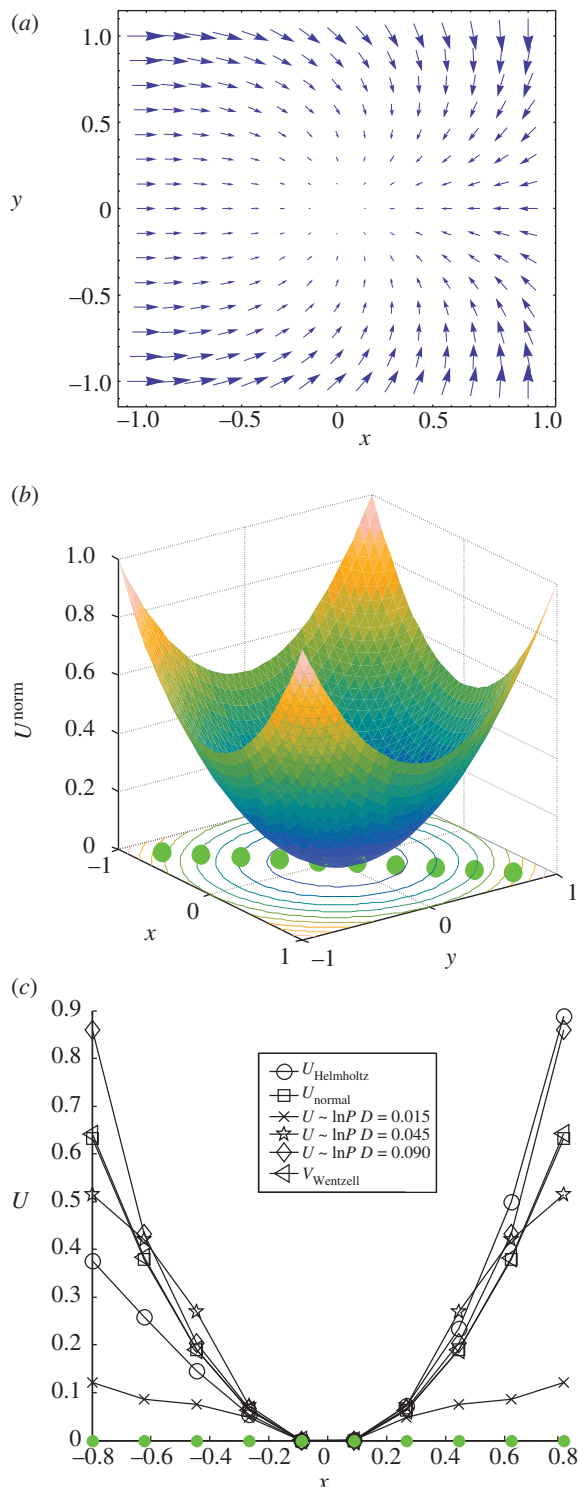
(*a*)

(*b*)

(*c*)

Figure 3. (*a*) Vector field of the driving forces of the dynamic system in example 1. Note how the non-vanishing remainder components $U_\perp$ cause the vector field to be non-symmetric even if the underlying quasi-potential U^{norm} is symmetric. (*b*) The quasi-potential U^{norm} reconstructed from the normal decomposition and the positions of nine points for comparison in (*c*). (*c*) Different quasi-potentials constructed from the Helmholtz decomposition, the normal decomposition, $-\ln P$ decomposition and the Freidlin–Wentzell action function for comparison of the nine positions (states) in phase space. (Online version in colour.)

0.015), U^{prob} (shown by the symbol 'x') does not 'cover' the peaks on both sides. Second, even if the noise level D is large enough to sample the whole landscape, its quasi-

potentials U^{prob} (shown by pentagons and diamonds) are not symmetrical owing to the asymmetric probability flux in the phase space. Here only the quasi-potential derived from normal decomposition U^{norm} (shown by squares) agrees with half of the values of the Freidlin–Wentzell potential (shown by triangles). The numerical method for calculating the Freidlin–Wentzell potential is in the electronic supplementary material.

5.2. Example 2. Two-dimensional dynamical system with four attractors

Because our original motivation was to determine the relative metastability of attractors in a multi-stable dynamical system, an example with four attractors is chosen to illustrate the various quasi-potential functions that can be extracted from the decomposition methods and allow us to compare the transition rates and global ordering of attractors they predict. (The Helmholtz decomposition is excluded from the analysis for its obvious invalidity.) Theoretical predictions from the two other decompositions are compared with the Freidlin–Wentzell action functions that can represent the transition rate between attractors caused by stochastic perturbation with large deviations.

The governing equations of our system are

$$
\left.
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \boldsymbol{F}_1(x,y) = -1 + 9x - 2x^3 + 9y - 2y^3 \\
\text{and} \quad \frac{\mathrm{d}y}{\mathrm{d}t} &= \boldsymbol{F}_2(x,y) = 1 - 11x + 2x^3 + 11y - 2y^3.
\end{aligned}
\right\}
$$
(5.7)

Because the dynamical system here is constructed to admit an analytical solution, it has a symmetric structure and can be decomposed exactly into a gradient part and a perpendicular remainder. However, the biological dynamic systems, such as the formulation of gene regulatory networks as rate equations for changes of gene expression, in general have no analytical solutions. Numerical methods are required to find the solutions approximately. (More details are included in the electronic supplementary material.)

$$
\left.
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \boldsymbol{F}_1(x,y) = -\frac{\partial U^{\text{norm}}}{\partial x} + \boldsymbol{F}_\perp^x = (10x - 2x^3 - y - 1) \\
&\qquad\qquad\qquad\qquad\qquad + (10x - 2y^3 - x) \\
\text{and} \quad \frac{\mathrm{d}y}{\mathrm{d}t} &= \boldsymbol{F}_2(x,y) = -\frac{\partial U^{\text{norm}}}{\partial y} + \boldsymbol{F}_\perp^y = (10y - 2y^3 - x) \\
&\qquad\qquad\qquad\qquad\qquad - (10x - 2y^3 - y - 1).
\end{aligned}
\right\}
$$
(5.8)

The quasi-potential function U^{norm} from the normal decomposition is shown in figure 4*a*. It has four attractors designated as A, B, C and D, with the quasi-potential function U^{norm} as follows:

$$
U^{\text{norm}}(x,y) = -5(x^2 + y^2) + \frac{1}{2}(x^4 + y^4) + xy + x. \quad (5.9)
$$

We can now examine the ordering of the attractors with respect to their relative metastability. Using equation (5.9), we find: $U^{\text{norm}}(D) > U^{\text{norm}}(B) > U^{\text{norm}}(C) >$
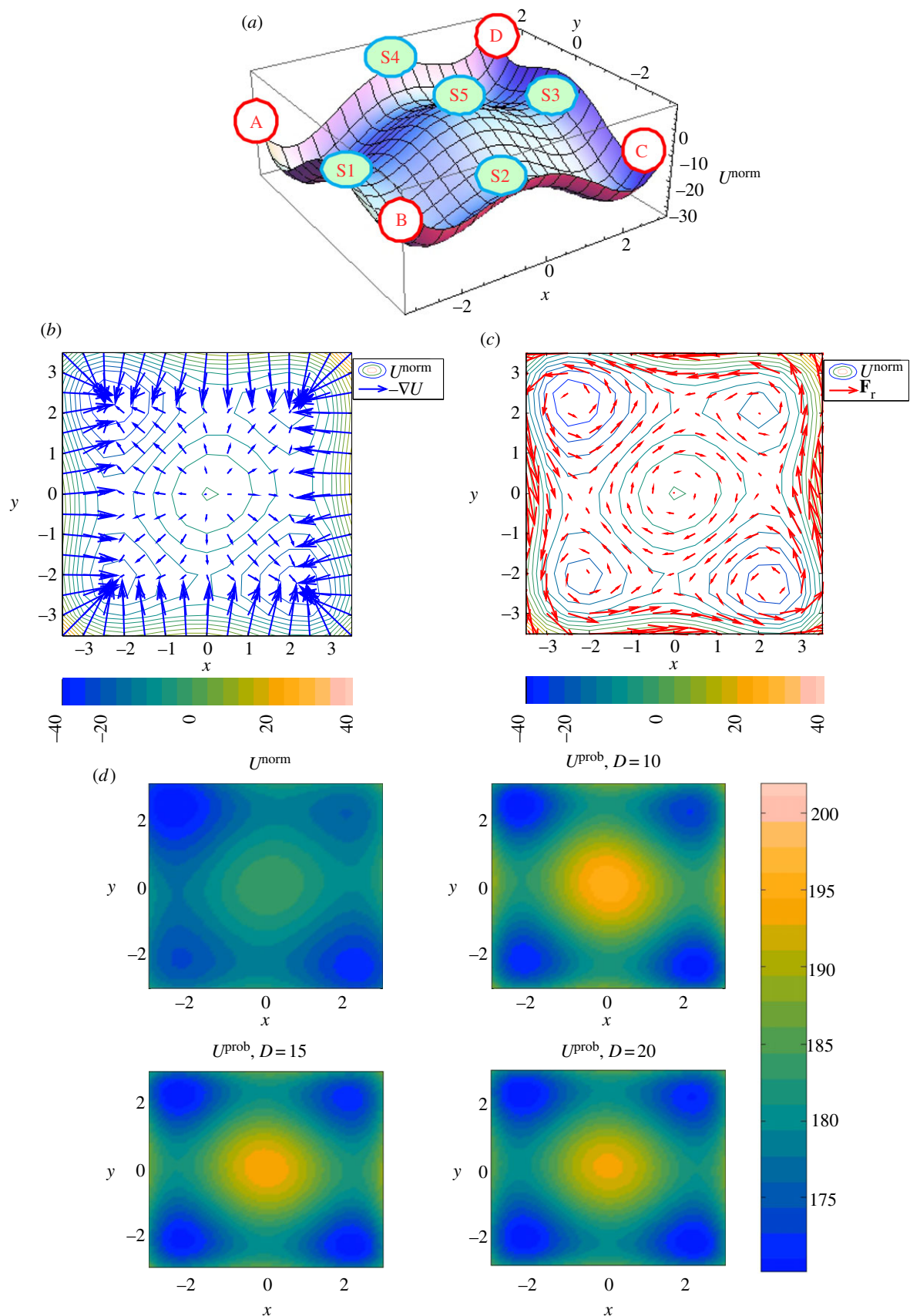
Figure 4. (*a*) Quasi-potential function U^{norm} derived from the normal decomposition. The white circles represent attractors A, B, C, D; the grey circles denote the saddle points. (*b*) Vector field plot of the gradient component of the normal decomposition super-positioned on the contour plot of U^{norm}. (*c*) Vector field plot of the $\boldsymbol{F}_{\perp}$, the remainder component of the normal decomposition super-positioned on the contour plot of U^{norm}. (*d*) Quasi-potential U^{norm} calculated from the normal decomposition and the U^{prob} calculated from the $-\ln P$ decomposition with various noise levels ($D = 10$, 15 and 20). (Online version in colour.)
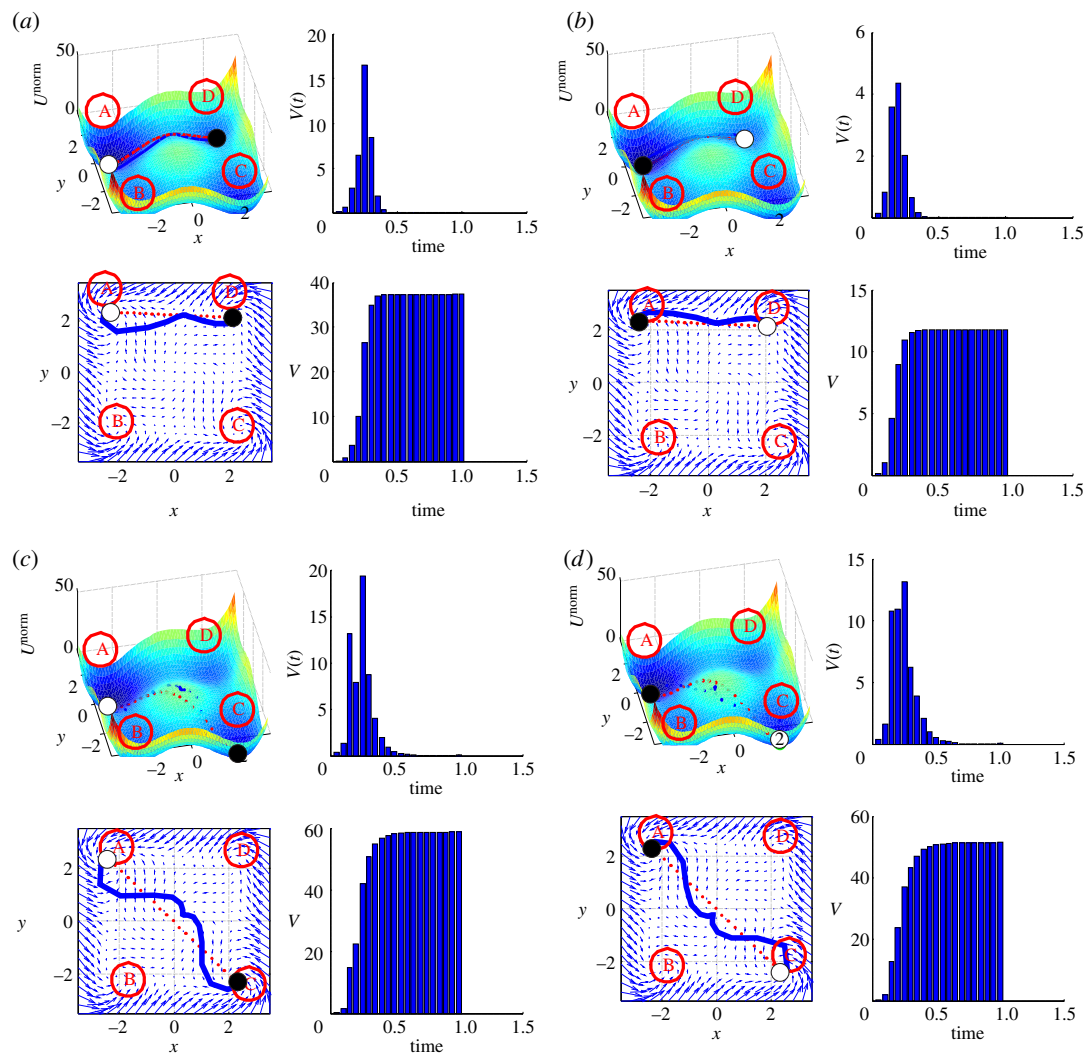
Figure 5. (*a*) The LAP for the attractor transition from x_A to x_D. The white dot denotes the starting point; the black dot is the end point. The attractors A, B, C, D are noted as in figure 4. $V(t)$ is the Wentzell action function at every time step, and V is the total Wentzell action function at each time t. (*b*) The LAP from attractor from x_D to x_A is different from the LAP from x_A to x_D. Here the Wentzell action function is smaller than for the transition from x_A to x_D because the attractor state x_D has a 'higher elevation' than x_A in the quasi-potential 'landscape' U^{norm}. (*c*) The LAP for the attractor transition from x_A to x_C. It is clear that the Wentzell action function applies during the 'uphill' process, while it is zero during the 'downhill' process. (*d*) The LAP for the attractor transition from x_C to x_A. The Wentzell action function is smaller than for the transition from x_A to x_C because attractor x_C is 'higher' than x_A in the quasi-potential U^{norm}. (Online version in colour.)

$U^{\text{norm}}(A)$. In this case, we have a transitive set, i.e. if $P_{D \to B} > P_{B \to D}$ and $P_{B \to C} > P_{C \to B}$ then $P_{D \to C} > P_{C \to D}$ and so forth. However, this does not need to be so in general (see §6, for more details).

We can verify that the quasi-potential function $U^{\text{norm}}(x,y)$ in equation (5.9) satisfies the HJE (3.18). From the vector field plot in figure 4, it is clear that the gradient components contribute to the driving force toward the attractors (figure 4*b*) while the remaining perpendicular components capture the rotational movement in the phase space (figure 4*c*). However, because the rotational driving forces are perpendicular to the gradient components, they do not contribute to the transition between the different attractors.

Using the same method as in the previous example, the decomposition for extracting U^{prob} can be numerically computed with varying noise levels. In figure 4*d*, the results for the quasi-potential U^{prob} with various noise levels are plotted alongside U^{norm} from the normal decomposition. As shown in the previous example, the accuracy of U^{prob} suffers if the noise level is too small to cover the whole landscape. In this case, we need $D = 20$ to get a reasonably accurate landscape.

Then we need to calculate the Wentzell action function V as the gold standard to validate two decompositions here. Because the Wentzell action function is usually too complicated to be found analytically, equation (4.4) can be re-written in discrete form to find an approximate solution (see the electronic supplementary material for more details). We can calculate the LAP for the attractor transition from x_A to x_D and vice versa, and for the attractor transition from x_A to x_C and vice versa (figure 5). It is clear that the Wentzell action function applies during the 'uphill' process while it is zero during the 'downhill' process (figure 5). The LAP from attractor from x_D to x_A is different from the LAP from x_A to x_D.

Here the Wentzell action function is smaller than for the transition from $\boldsymbol{x}_A$ to $\boldsymbol{x}_D$ because the attractor state $\boldsymbol{x}_D$ has a 'higher elevation' than $\boldsymbol{x}_A$ in the quasi-potential 'landscape' U^{norm} (figure 5a,b). Note that the LAP $\mathbf{x_A} \rightarrow \mathbf{x_D}$ is also different from the reverse path $\mathbf{x_D} \rightarrow \mathbf{x_A}$ and the LAP $\mathbf{x_A} \rightarrow \mathbf{x_C}$ is different from the reverse path $\mathbf{x_C} \rightarrow \mathbf{x_A}$ (figure 5). Such 'path irreversibility' is a common characteristic of nonlinear dynamic systems.

On the basis of the Freidlin–Wentzell large deviation theory, the transition rate between two points can be determined from the transition barriers ΔU along the LAP. Instead of comparing the transition rates directly, ΔU calculated from the quasi-potential U^{norm} and by U^{prob} are compared with the Freidlin–Wentzell action functions V, as shown in figure 6. It is shown that U^{norm} agree well with the half of the Freidlin–Wentzell action functions V while U^{prob} mostly do not agree with it. It could be argued that the accuracy of U^{prob} could be further improved by using bigger noise and better numerical methods. However, from various numerical test, U_{prob} does not significantly improve its accuracy more than the example presented here.

6. DISCUSSION

We show here that the quasi-potential U^{norm} obtained from the normal decomposition has a non-local meaning, in that the sign of ΔU^{norm} represents the relative stability of a pair of adjacent attractors with respect to the spontaneity to transition from one attractor to another. The spontaneity implies that the system can, in the presence of noise in the state variables $\boldsymbol{x}$, i.e. without an explicit external driving force that directs this transition as epitomized by signalling molecules in cell biology, move from a given state to the most probable neighbouring state. No matter what the transition path is between two neighbouring attractors, $\mathbf{x_A}$ and $\mathbf{x_B}$, if $U_A^{\mathrm{norm}} > U_B^{\mathrm{norm}}$ it is always easier to transition from point $\mathbf{x_A}$ to point $\mathbf{x_B}$ and vice versa. By contrast, for the alternative decomposition based on $U^{\mathrm{prob}} \sim -\ln P$, this equivalency between potential and transition rate is not exact although, for practical purposes, $U_S^{\mathrm{prob}} - U_A^{\mathrm{prob}}$ (where U_s is the potential at the 'saddle' (exit) point between these attractors) often scales with $P_{\mathbf{x_A} \rightarrow \mathbf{x_B}}$ [16].

For the transition $\mathbf{x_A} \rightarrow \mathbf{x_C}$ in cases where the attractors $\mathbf{x_A}$ and $\mathbf{x_C}$ are not adjacent, but separated by attractor B that needs to be traversed, i.e. $\mathbf{x_A} \rightarrow \mathbf{x_B} \rightarrow \mathbf{x_C}$ with 'saddle' points x_{S_1} between $\mathbf{x_A}$, $\mathbf{x_B}$ and saddle points x_{S_2} between $\mathbf{x_A}$, $\mathbf{x_C}$, the above approach can be expanded [25],

$$P_{x_A \rightarrow x_C} = P_{x_A \rightarrow x_B} \cdot P_{x_B \rightarrow x_C} = \mathrm{e}^{-vAB/\varepsilon^2} \cdot \mathrm{e}^{-vBC/\varepsilon^2}$$
$$= \mathrm{e}^{-(U_{S_1}^{\mathrm{norm}} - U_A^{\mathrm{norm}})/\varepsilon^2} \cdot \mathrm{e}^{-(U_{S_2}^{\mathrm{norm}} - U_B^{\mathrm{norm}})/\varepsilon^2}. \quad (6.1)$$

The case with more than one attractor in between can also be calculated with a similar approach to that in equation (6.1).

Theories of quasi-potential functions have been developed for the one-to-one relationship of two points (Freidlin–Wentzell theory, Kramers' law). Two caveats are in order here regarding the interpretation of ΔU^{norm} in systems with more than two attractors.

6.1. ΔU *and 'fate choice' in multi-potent cells*

First, the attractors with lower quasi-potentials U^{norm} have higher probability to be occupied when a system reaches its steady state—which may take a long time for a rugged landscape with small noise. This shall remind us that ΔU^{norm} alone does not allow one to predict the 'fate choice' of a stem cell in an attractor within a multi-attractor system that faces alternative transitions (representing its potential to differentiate into multiple distinct cell types spontaneously). The local landscape topography needs to be considered. Similar to the case for equilibrium systems, speaking of potentials, irrespective of whether path-dependent or not, one needs to bear in mind that ΔU^{norm} between two points is not a general measure of spontaneity of a process but that the latter depends on the local kinetic barriers, as illustrated in figure 1. Referring to the constellation in figure 1 it is clear that noise (or temperature) may be just high enough but not too high, such that one will find that the system in attractor A will prefer to move to B rather than to C (in the absence of catalysis). Only under sufficient noise (or high temperature) and time will the preference for C in steady state become manifest.

6.2. *A global potential landscape?*

Second, we now return to the more profound motivation behind the interest for a global quasi-potential function: the assessment of the 'relative (meta)stability' of attractor states, that is, their ordering with respect to spontaneous transitions into each other. Such a global behaviour of a system, as opposed to local, peri-attractor behaviour, is the scale of system change at which complex biological processes, such as cell differentiation during the development, take place. But is such an ordering possible? Can we directly compare the values of the quasi-potentials $U_i^{\mathrm{norm}}(i = 1, 2, 3, \ldots, m)$ for m attractors and extract meaningful information as to long-term fate choice—aside from the aforementioned issue of kinetic inhibition? This is tempting to ask because, unlike the Freidlin–Wentzell potential difference ΔV, which is defined for pairs of states, we can compute from the system equations directly the potential value $U^{\mathrm{norm}}(S)$ for any state S. If U^{norm} captures the global behaviour, one would postulate that our conjectured third condition, the transitivity of any three points a, b, c with respect to their quasi-potential U, would be satisfied: if $P_{a \rightarrow b} > P_{b \rightarrow a}$ and $P_{b \rightarrow c} > P_{c \rightarrow b}$, then $P_{a \rightarrow c} > P_{c \rightarrow a}$. Example 2 constitutes such a transitive system. However, there is no guarantee that transitivity is satisfied for all systems. Although ΔU^{norm} scales monotonically with ΔV (Wentzell action function), when it is defined for points within the same basin of attraction (most usefully, as the difference between the attractor state and the exit state at the saddle), there is no simple mathematical relationship between ΔU^{norm} and ΔV along the transition trajectory between two attractors and it thus is 'agnostic' for global ordering.

Only in those cases where there is transitivity among all the potential values of attractors i, U_i^{norm}
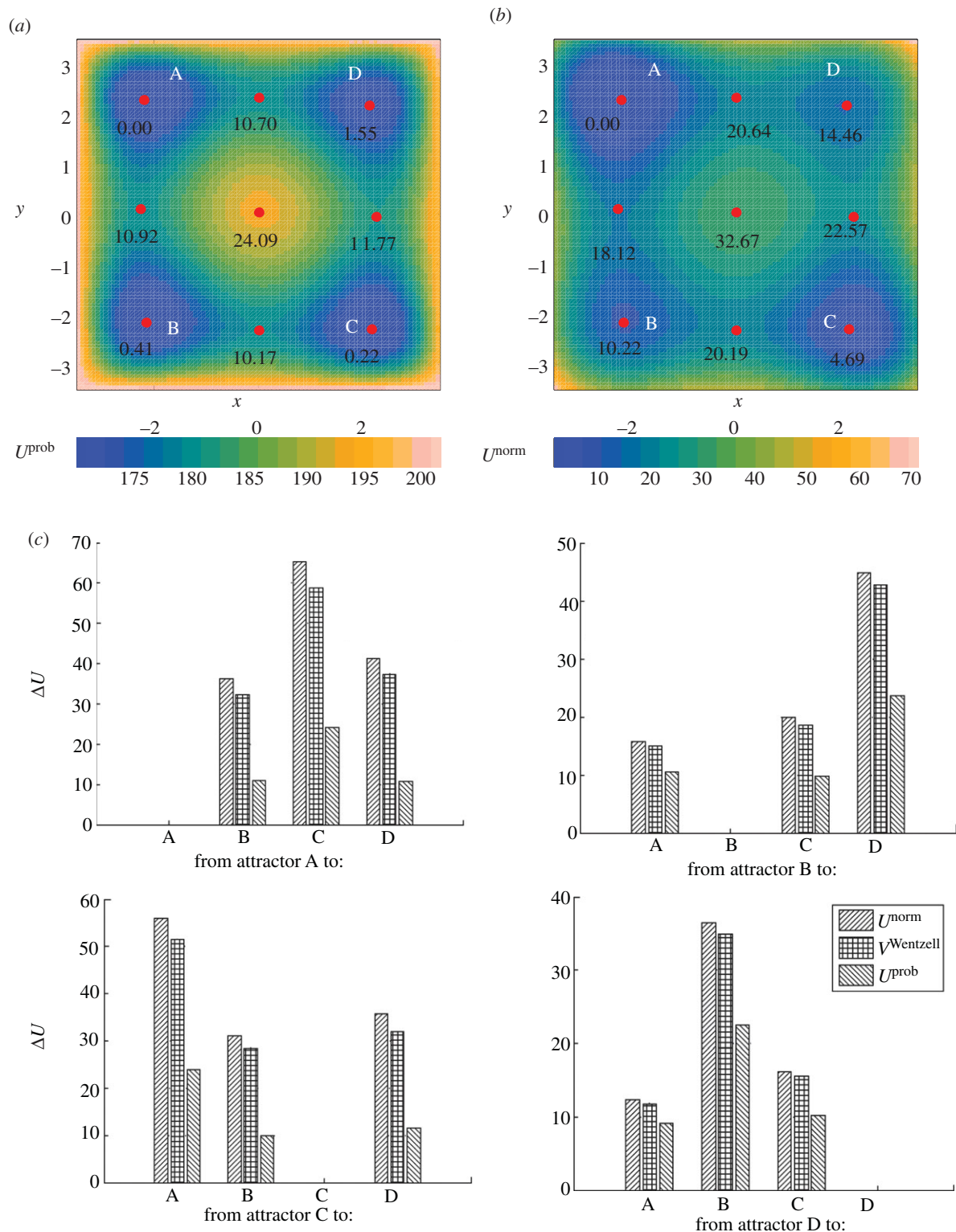
Figure 6. (*a*) Quasi-potential U^{prob} at the attractors and 'saddle' points derived from the $-\ln P$ decomposition with noise level $D = 20$. (*b*) Quasi-potential U^{norm} at the same points derived from the normal decomposition. Attractors A, B, C, D are denoted as in figure 4. (*c*) Potential barriers ΔU that need to be overcome for the transition from each attractor to every other one. ΔU were calculated using the quasi-potential function from the normal decomposition U^{norm} and $-\ln P$ decomposition U^{prob} based on the topology of attractors and 'saddle' points in between. Then ΔU were compared with the Freidlin–Wentzell action functions V calculated from numeric minimization. (Online version in colour.)

(consistency) in the directed graph in which the edge arrows depict the direction of higher transition probability $P(\mathbf{x}_i \rightarrow \mathbf{x}_j)$ between pairs of attractors $\mathbf{x}_i$ and $\mathbf{x}_j$, can we speak of a 'global' landscape. This

condition surely will not be satisfied for systems exhibiting biological phenomena such as circadian cycle, cell cycle and other biological limit cycles. However, in some classes of systems, as lucidly epitomized

in the qualitative picture of Waddington's epigenetic landscape that describes the essential features of cell development or in the fitness landscape of evolution, such global landscapes appear to adequately capture the developmental or evolutionary dynamics and constraints. In differentiation, it visualizes the barriers of trans-differentiation or reprogramming from one cell type to another. This suggests the possibility of a global ordering of (meta)stability for a particular yet to be defined subclass of networks to which our example 2 belongs, in which the global transitivity of U is satisfied. Thus, it could be that evolved networks that govern acyclic phenotype transitions, such as development or evolution, have for some unknown reasons a quasi-potential landscape that satisfies the transitivity condition (iii)—which is a subset of multi-attractor systems. The facility to compute for each attractor i of a biological network a quasi-potential U_i opens the possibility to study this phenomenon.

6.3. Present and future biological significance

Knowledge of the quasi-potential U^{norm}, if it can be computed, has direct significance for understanding and manipulating regulation of developmental paths in multicellular systems. This would require an accurate dynamic system description of the relevant gene regulatory network and measurement of the state vector $\boldsymbol{x}_i^*$ for an attractor state of interest—corresponding to biologically defined cell types. $U^{\mathrm{norm}}(\boldsymbol{x}_i^*)$ can then be computed from the normal decomposition of the system equations for the attractor states $\boldsymbol{x}_i^*$, which currently can be approximately measured as transcriptome or proteome. $U^{\mathrm{norm}}(\boldsymbol{x}_i^*)$ in turn will afford a 'global ordering' of the cell types i with regard to their metastability. The ordering of a set of different cell types can be used to predict their relative stability in terms of the 'ease' (probability) of transitions into 'neighbouring' cell types, i.e. transitions between attractor states under noise or perturbation. The gradient of U^{norm} can be interpreted as the driving force that imposes directionality to the process of cell differentiation. This ultimately can help explain how the gene–gene interactions encoded in the genomic DNA sequence unfold to manifest as a robust, spontaneous and directed developmental process.

More concretely, because U^{norm} of two different cell types also provide an objective measure to compare their relative (meta)stability, we could in principle find the best subset of genes whose expression needs to be jointly perturbed to minimize the 'transition barrier' between these two cellular attractor states. This could pave the way towards a rational design of cell reprogramming protocols for which it is currently only known that combinations of a number of genes need to be perturbed, the determination of which however is purely empirical.

At present, the major limitation to realizing the above utility of the quasi-potential formalism is that information about the structure of the gene regulatory network, $\boldsymbol{F}(\mathrm{x})$, is only sketchy at best, limiting its description as a dynamic system (equation (1.1))—the

prerequisite for determining U^{norm} (x). As we learn more details about $\boldsymbol{F}(x)$ the quasi-potential will gain in importance accordingly.

REFERENCES

1 Brock, A., Chang, H. & Huang, S. 2009 Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genetics* **10**, 336–342. (doi:10.1038/nrg2556)

2 Huang, S. & Kauffman, S. 2009 Complex gene regulatory networks—from structure to biological observables: cell fate determination. In *Encyclopedia of complexity and systems science* (ed. R. A. Meyers). Berlin, Germany: Springer.

3 Nicolis, G. 1986 Dissipative systems. *Rep. Prog. Phys.* **49**, 873–949. (doi:10.1088/0034-4885/49/8/002)

4 Haenggi, P., Talkner, P. & Borkovec, M. 1990 Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341. (doi:10.1103/RevModPhys.62.251)

5 Freidlin, M. I. & Wentzell, A. D. 1984 *Random perturbations of dynamical systems*. Berlin, Germany: Springer.

6 Lyapunov, A. M. 1892 The general problem of the stability of motion. *Int. J. Control* **55**, 531–534.

7 Prigogine, I. 1997 *The end of certainty 1*. New York, NY: The Free Press.

8 Efimov, D. 2009 On global Lyapunov characterization of multi-stable nonlinear systems. In *Proc. 48th IEEE Conf. on Decision and Control, 2009, held jointly with the 2009 28th Chinese Control Conf., Shanghai, People's republic of China, 15–18 December 2009,* pp. 6299–6304. (doi:10.1109/CDC.2009.5399944)

9 Bhattacharya, S., Zhang, Q. & Andersen, M. E. 2011 A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst. Biol.* **5**, 85. (doi:10.1186/1752-0509-5-85)

10 Yuan, R., Yuan, B. & Ao, P. 2010 Constructive proof of global Lyapunov function as potential function. (http://arxiv.org/abs/1012.2721v1)

11 Ao, P. 2009 Global view of bionetwork dynamics: adaptive landscape. *J. Genet. Genomics* **36**, 63–73. (doi:10.1016/S1673-8527(08)60093-4)

12 Ao, P., Galas, D., Hood, L. & Zhu, X. 2008 Cancer as robust intrinsic state of endogenous molecular-cellular network shaped by evolution. *Med. Hypotheses* **70**, 678–684. (doi:10.1016/j.mehy.2007.03.043)

13 Ao, P. 2002 Stochastic force defined evolution in dynamical systems. (http://arxiv.org/abs/physics/0302081v1)

14 Kwon, C., Ao, P. & Thouless, D. J. 2005 Structure of stochastic dynamics near fixed points. *Proc. Natl Acad. Sci.* **102**, 13 029–13 033. (doi:10.1073/pnas.0506347102)

15 Xing, J. 2010 Mapping between dissipative and Hamiltonian systems. *J. Phys. A Math. Theoret.* **43**, 375003. (doi:10.1088/1751-8113/43/37/375003)

16 Wang, J., Xu, L., Wang, E. & Huang, S. 2010 The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.* **99**, 29–39. (doi:10.1016/j.bpj.2010.03.058)

17 Berglund, N. 2011 Kramers' law: Validity, derivations and generalisations. (http://arxiv.org/abs/1106.5799)

18 Griffiths, P. & Harris, J. 1994 *Principles of algebraic geometry*. New York, NY: Wiley-Interscience.

19 Li, C., Wang, E. & Wang, J. 2011 Potential landscape and probabilistic flux of a predator prey network. *PLoS ONE.* **6**, e17888. (doi:10.1371/journal.pone.0017888)

20 Micheelsen, M. A., Mitarai, N., Sneppen, K. & Dodd, I. B. 2010 Theory for stability and regulation of epigenetic landscapes. *Phys. Biol.* **7**, 026010. (doi:10.1088/1478-3975/7/2/026010)

21 Wang, J., Xu, L. & Wang, E. 2008 Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl Acad. Sci. USA* **105**, 12 271–12 276. (doi:10.1073/pnas.0800579105)

22 Wang, J., Zhang, K., Xu, L. & Wang, E. 2011 Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl Acad. Sci. USA* **108**, 8257–82562. (doi:10.1073/pnas.1017017108)

23 Ao, P., Kwon, C. & Qian, H. 2007 On the existence of potential landscape in the evolution of complex systems. *Complexity* **12**, 19–27. (doi:10.1002/cplx.v12:4)

24 Aurell, E. & Sneppen, K. 2002 Epigenetics as a first exit problem. *Phys. Rev. Lett.* **88**, 2–5. (doi:10.1103/PhysRevLett.88.048101)

25 Maier, R. S. & Stein, D. L. 1994 Asymptotic exit location distributions in the stochastic exit problem. *SIAM J. Appl. Math.* **57**, 752–790. (doi:10.1137/S0036139994271753)