Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory*

Simone Bruno[†,‡], Felipe A. Campos[†,§], Yi Fu[§], Domitilla Del Vecchio[‡], Ruth J. Williams[§]

Abstract.
Epigenetic cell memory, the inheritance of gene expression patterns across subsequent cell divisions, is a critical property of multi-cellular organisms. In recent work [10], a subset of the authors observed in a simulation study how the stochastic dynamics and time-scale differences between establishment and erasure processes in chromatin modifications (such as histone modifications and DNA methylation) can have a critical effect on epigenetic cell memory. In this paper, we provide a mathematical framework to rigorously validate and extend beyond these computational findings. Viewing our stochastic model of a chromatin modification circuit as a singularly perturbed, finite state, continuous time Markov chain, we extend beyond existing theory in order to characterize the leading coefficients in the series expansions of stationary distributions and mean first passage times. In particular, we characterize the limiting stationary distribution in terms of a reduced Markov chain, provide an algorithm to determine the orders of the poles of mean first passage times, and determine how changing erasure rates affects system behavior. The theoretical tools developed in this paper not only allow us to set a rigorous mathematical basis for the computational findings of our prior work, highlighting the effect of chromatin modification dynamics on epigenetic cell memory, but they can also be applied to other singularly perturbed Markov chains beyond the applications in this paper, especially those associated with chemical reaction networks.

Key words. singular perturbation, continuous time Markov chain, multimodal stationary distribution, mean first passage times, epigenetic cell memory, chromatin modification circuits

MSC codes. 92C40, 92C42, 60J28

1 Introduction

1.1 Background Epigenetic cell memory, the inheritance of gene expression patterns across subsequent cell divisions [22], is a critical property of multi-cellular organisms of intense interest in the field of systems biology [30, 31]. It has previously been discovered that chromatin modifications, such as DNA methylation and histone modifications, are key mediators of epigenetic cell memory [1, 14, 21, 24] (see references in [10] for more biological background). More precisely, it was found via simulations of stochastic models that the time scale separation between establishment (fast) and erasure (slow) of these modifications extends the duration of cell memory, and that different asymmetries between erasure rates of chromatin modifications can lead to different gene expression patterns [10–12]. Here, we provide a mathematical framework to rigorously validate these computational findings and to further explore models

[†]These authors contributed equally: S. Bruno and F. A. Campos.

[‡]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 (sbruno@mit.edu, ddv@mit.edu).

[§]Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0112 (fcamposv@ucsd.edu,yif064@ucsd.edu,rjwilliams@ucsd.edu).

*Funding. S.B. was supported by NSF Collaborative Research grant MCB-2027949 (PI: D.D.V.). R.J.W., F.C and Y.F. were supported in part by NSF Collaborative Research grant MCB-2027947 (PI: R.J.W.) and by the Charles Lee Powell Foundation (PI: R.J.W.).

of chromatin modification circuits. We do this in a way that the results obtained and the tools developed can be applied to other mathematical models beyond the applications in this paper, especially stochastic models of chemical reaction networks.

1.2 Focus of our work In this paper, we consider different versions of the chromatin modification circuit proposed in [10]. In particular, we start with simpler circuits that include histone modifications only and then we consider more elaborate circuits that include also DNA methylation. All of these circuits can be viewed as examples of Stochastic Chemical Reaction Networks (SCRNs). A SCRN is a continuous time Markov chain living in the non-negative integer lattice in d-dimensions, where the components of the Markov chain track the number of molecules of each of d species in the network over time, and each jump of the Markov chain corresponds to the firing of a reaction in the network [2]. A more precise description is given in Section 3.2.

In order to analyze these stochastic models, we first determine how the stationary distributions and mean first passage times between states vary when a small parameter ε (non-dimensional parameter that scales the speed of the basal erasure of all the chromatin modifications) tends to zero. To this end, we show that the stationary distributions and the mean first passage times of these singularly perturbed Markov chains admit series expansions in ε and we develop theoretical tools to determine the coefficients in these expansions. Then, we focus on determining how the different erasure rates of chromatin modifications affect the behavior of the chromatin modification circuit models. This latter study is conducted by exploiting comparison theorems for Markov chains recently developed in [13].

One of the key features of our work is that these tools and the associated mathematical results are not only applicable to the chromatin modification models, but they can also be used to analyze other models that respect the same set of assumptions.

1.3 Related work As mentioned in the previous paragraph, the stochastic behavior of the chromatin modification circuit models can be described by singularly perturbed continuous time Markov chains. There is some literature on discrete and continuous time, singularly perturbed Markov chains, especially by Avrachenkov et al. [6], Hassin & Haviv [20], Beltrán and Landim [7,8], and Yin & Zhang [32]. Avrachenkov et al. [6] gave general characterizations of series expansions for the stationary distribution and mean first passage times of a singularly perturbed *discrete time* Markov chain with finite state space. While their theory can be in principle translated to continuous time Markov chains, our work mostly deals directly with the singularly perturbed continuous time Markov chains and provides more concrete theoretical results for the leading coefficients of the stationary distribution series expansion and the orders of the poles of the mean first passage times. For the leading coefficients in the series expansion for the mean first passage times, we use in part the results of Avrachenkov & Haviv [5] and Avrachenkov et al. [6] and adapt their work to the continuous time Markov chain setting. We treat in detail the case where the chain for $\varepsilon = 0$ has more than one absorbing state and at least one transient state. Furthermore, we also provide an interpretation of leading coefficients in the series expansion of the stationary distribution in terms of a certain restricted Markov chain. An algorithm we give to determine the order of the pole of the mean first passage time extends the work of Hassin & Haviv [20] from discrete time to continuous time. We also extend the original algorithm's scope to treat mean first passage times to a subset of states, instead of just a single state. Beltrán and Landim [7,8] study metastable and tunneling behavior for a sequence $\{\eta^N\}_{N=1}^{\infty}$ of time-homogeneous continuous time Markov chains with countable state spaces. Under an acceleration of time by a factor θ_N, they give conditions under which the trace of the accelerated process on the metastates is asymptotically Markovian as $N \to \infty$. For our case, this would correspond to accelerating time for $\eta^N = X^\varepsilon$ by $\theta_N \approx \frac{1}{\varepsilon}$. Beltrán and Landim identified the transition rates for the limiting Markov chain and proved that its stationary distribution can be obtained as a limit from the stationary distribution for η^N. While this work is potentially related to what we did, it requires knowing the stationary distribution for η^N *a priori*. Our approach does not need to know that stationary distribution explicitly and we also study mean first passage times, giving explicit asymptotics for both. Finally, Yin & Zhang [32] conducted an extensive study focused on determining matched asymptotic expansions for the marginal distributions at time t of singularly perturbed continuous time Markov chains. Their infinitesimal generators, generalizing those of Phillips & Kokotovic [29] and Pan & Basar [28], are of the form $Q(\varepsilon) = \frac{1}{\varepsilon}Q^{(0)} + Q^{(1)}$, and can be time dependent. For the time independent case, this would correspond to studying the marginal distributions of our Markov chain X^ε in the "linear" case and at time $\frac{t}{\varepsilon}$ as $\varepsilon \to 0$, i.e., $\lim_{\varepsilon\to 0} X^\varepsilon(\frac{t}{\varepsilon})$. Thus, while their work potentially might provide information about stationary distributions as $\varepsilon \to 0$, we directly study the power series expansion (in ε) of the stationary distribution of X^ε, and we also study series expansions of mean first passage times for X^ε, and we develop more concrete analyses for both.

1.4 Outline of the paper In Section 2 we introduce two simplified models for the chromatin modification circuit that do not include DNA methylation. Through these examples, we introduce the mathematical setting and questions we address in this paper. We describe the basic setup and definitions needed for this paper in Section 3. We present our main results in Section 4. Some proofs are given there, whilst others are in the Supplementary Information (SI). Further applications of the theoretical tools developed in Section 4 for chromatin modification circuits that include DNA methylation are presented in Section 5. Concluding remarks are given in Section 6.

1.5 Preliminaries and notation Denote the set of integers by $\mathbb{Z}$. For an integer $d \geq 2$ we denote by $\mathbb{Z}^d$ the set of d-dimensional vectors with entries in $\mathbb{Z}$. Denote by $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$, the set of non-negative integers. For an integer $d \geq 2$ we denote by $\mathbb{Z}_+^d$ the set of d-dimensional vectors with entries in $\mathbb{Z}_+$. We denote by $\mathbb{1}$ a vector of any dimension where all entries are 1's. The size of $\mathbb{1}$ will be understood from the context. The set of real numbers will be denoted by $\mathbb{R}$, $\mathbb{R}_+ = [0, \infty)$, $\mathbb{R}_{>0} = (0, \infty)$, and d-dimensional Euclidean space will be denoted by $\mathbb{R}^d$ for $d \geq 2$. For integers $n, m \geq 1$, the set of $n \times m$ matrices with real-valued entries will be denoted by $\mathbb{R}^{n\times m}$. The set of complex numbers will be denoted by $\mathbb{C}$.

Let $\mathcal{X}$ be a finite set. If needed, we will enumerate the entries of $\mathcal{X}$ by $\{1, \ldots, |\mathcal{X}|\}$. For a matrix $A = (A_{x,y})_{x,y\in\mathcal{X}}$ with real-valued entries, we denote the kernel of A by $\ker(A) := \{x \in \mathbb{R}^{|\mathcal{X}|} : Ax = 0\}$ and the nullity of A by $\text{nullity}(A) := \dim(\ker(A))$. We denote the spectrum of A by $\text{sp}(A)$ and the spectral radius by $\text{spr}(A) = \max\{|\lambda| : \lambda \in \text{sp}(A)\}$. A matrix $Q = (Q_{x,y})_{x,y\in\mathcal{X}}$ will be called a Q-matrix if $Q_{x,y} \geq 0$ for every $x \neq y \in \mathcal{X}$ and $Q\mathbb{1} = 0$. We denote the identity matrix, which has 1's on the diagonal and zeros elsewhere, by $I = (I_{x,y})_{x,y\in\mathcal{X}}$. For a vector $v = (v_x)_{x\in\mathcal{X}}$ we denote by $\text{diag}((v_x)_{x\in\mathcal{X}})$ the diagonal matrix in $\mathcal{X}$ with entries given by v. Vectors are column vectors unless indicated otherwise and a superscript of T will denote the transpose of a vector or matrix. For integers $n, m \geq 1$ and a matrix $A \in \mathbb{R}^{n\times m}$, we denote by $\|A\| = (\sum_{i=1}^{n}\sum_{j=1}^{m}|A_{i,j}|^2)^{1/2}$ the Frobenius norm of A[1]. For a vector $v \in \mathbb{R}^n$, we denote the Euclidean norm of v by $\|v\| = (\sum_{i=1}^{n}|v_i|^2)^{1/2}$.

Definition 1.1. *Given a matrix $A^{(0)}$ in $\mathbb{R}^{n\times m}$, a* ***real-analytic perturbation*** *of $A^{(0)}$ is a matrix-valued function $A : [0,\varepsilon_0) \longrightarrow \mathbb{R}^{n\times m}$, where $\varepsilon_0 > 0$, and*

$$A(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k A^{(k)}, \qquad 0 \leq \varepsilon < \varepsilon_0, \tag{1.1}$$

in which $\{A^{(k)} : k \geq 0\}$ is a sequence of matrices in $\mathbb{R}^{n\times m}$ such that

$$\sum_{k=0}^{\infty} \varepsilon^k \|A^{(k)}\| < \infty, \qquad \text{for every } 0 \leq \varepsilon < \varepsilon_0. \tag{1.2}$$

Such a perturbation is called ***linear*** *if $A(\varepsilon) = A^{(0)} + \varepsilon A^{(1)}$ for $0 \leq \varepsilon < \varepsilon_0$.*

By (1.2), a real-analytic perturbation of $A^{(0)}$ can be extended to a function $F(z) := \sum_{k=0}^{\infty} z^k A^{(k)}$ defined on $B(0,\varepsilon_0) = \{z \in \mathbb{C} : |z| < \varepsilon_0\}$. The function F will be called an **analytic perturbation** or **complex-analytic perturbation** of $A^{(0)}$. This extension will allow us to invoke results in complex analysis in order to study real-analytic perturbations. An example of this is the following result.

Proposition 1.2. *Let $A : [0,\varepsilon_0) \longrightarrow \mathbb{R}^{n\times n}$ be a real-analytic perturbation of $A^{(0)}$ such that $A^{-1}(\varepsilon)$ exists for every $0 < \varepsilon < \varepsilon_0$. Then, there is $\varepsilon_1 \in (0,\varepsilon_0)$ and $p \in \mathbb{Z}_+$ such that*

$$A^{-1}(\varepsilon) = \sum_{k=-p}^{\infty} \varepsilon^k B^{(k)}, \quad 0 < \varepsilon < \varepsilon_1, \tag{1.3}$$

where $\sum_{k=-p}^{\infty} \varepsilon^k \|B^{(k)}\| < \infty$ for every $0 < \varepsilon < \varepsilon_1$, $\{B^{(k)} : k \geq -p\}$ is a sequence of matrices in $\mathbb{R}^{n\times n}$, $B^{(-p)}$ is not the identically zero matrix and p is called the ***order of the pole*** *at $\varepsilon = 0$.*

This result is given in the analytic setting as Theorem 2.4 in [6]. Proposition 1.2 follows by extending $A(\cdot)$ to a complex disk, then using Cramer's rule as in the proof of Theorem 2.4 in [6] and checking that the matrices $\{B^{(k)} : k \geq -p\}$ obtained are real-valued.

2 Motivating Example: Chromatin Modification Circuit

In order to understand how the interactions among known chromatin modifications influence epigenetic cell memory, we consider the chemical reaction model of the gene's inner chromatin modification circuit introduced in [10]. This model has the nucleosome with DNA wrapped around it, D, as a basic unit that can be modified either with activating marks, such as H3K4 methylation (H3K4me3) or H3K4 acetylation (H3K4ac), or repressive marks, such as H3K9 methylation (H3K9me3) or DNA methylation. H3K4me3 and H3K4ac are two histone modifications that promote a less

[1]Here, we chose to fix a particular norm on $\mathbb{R}^{n\times m}$, although other choices of norm will often work.

compact DNA around the nucleosomes and they are then associated with gene activation (see Chapter 3 of [1] and [33]). In the model, it is assumed that H3K4me3 and H3K4ac co-exist and the nucleosome with either of these modifications is represented by $\mathrm{D^A}$. On the contrary, both the histone modification H3K9me3 and DNA methylation cause the DNA to be tightly wrapped around the nucleosome and therefore, they are associated with gene repression [22]. A nucleosome with DNA methylation only, H3K9 methylation (H3K9me3) only or both is represented by $\mathrm{D^R_1}$, $\mathrm{D^R_2}$ and $\mathrm{D^R_{12}}$, respectively.

One of the key parameters of the system is $\varepsilon > 0$, a non-dimensional parameter that scales the speed of basal erasure of all chromatin modifications. We are interested in studying the behavior of the system in the limiting regime $\varepsilon \to 0$, in which the chromatin modification system has a bimodal limiting stationary distribution [10]. One peak corresponds to the active chromatin state (most of the nucleosomes are modified with activating marks) and the other one is in the repressed chromatin state (most of the nucleosomes are modified with repressive marks). We aim to derive formulas that characterize, as ε goes to 0, the behavior of the stationary distribution and the "time to memory loss" of the active (repressed) state, defined as the mean first passage time to reach the repressed (active) state, starting from the active (repressed) state.

Two other critical parameters of the system are μ and μ': they capture the asymmetry between the erasure rates of repressive and activating chromatin modifications. More precisely, μ (μ') quantifies the asymmetry between erasure rates of repressive histone modifications (DNA methylation) and activating histone modifications. Part of our study is to analytically determine how μ and μ' affect the stationary distribution and the time to memory loss of the active and repressed states.

In this section, we introduce two simplified models of the chromatin modification circuit in which, compared to the full model described above, DNA methylation is not included and the only chromatin marks are histone modifications. We will use these simpler models in Section 4 to directly apply and then better understand the theory developed in this paper. Then, in Section 5 we deal with more elaborate models that also include DNA methylation. Note that, for consistency, we use the same notation for the species and the reaction rate constants as the one used in the paper where these models were introduced [10].

2.1 1D model We first consider a simplified model in which a gene has a total of $\mathrm{D_{tot}} \geq 2$ nucleosomes, where each nucleosome either has an activating histone modification, $\mathrm{D^A}$, or a repressive histone modification, $\mathrm{D^R}$, and there are no unmodified nucleosomes in this simplified model. If the amounts of nucleosomes having repressive ($\mathrm{D^R}$) and activating ($\mathrm{D^A}$) modifications are denoted as $n_{\mathrm{D^R}}$ and $n_{\mathrm{D^A}}$, respectively, then we have the conservation law $n_{\mathrm{D^R}} + n_{\mathrm{D^A}} = \mathrm{D_{tot}}$. We call this the 1D model because it suffices to keep track of the amount of $\mathrm{D^R}$ (for example), since the amount of $\mathrm{D^A}$ can be deduced by the conservation law. Furthermore, the basal and recruited erasure of $\mathrm{D^A}$ ($\mathrm{D^R}$) coincide with the basal *de-novo* establishment and maintenance of $\mathrm{D^R}$ ($\mathrm{D^A}$). The chemical reaction system for this 1D model is the following:

$$\text{(2.1)} \qquad \begin{aligned} &① \ \mathrm{D^A} + \mathrm{D^R} \xrightarrow{k_E^A} \mathrm{D^R} + \mathrm{D^R}, \quad ② \ \mathrm{D^A} \xrightarrow{\delta + \bar{k}_E^A} \mathrm{D^R}, \\ &③ \ \mathrm{D^R} + \mathrm{D^A} \xrightarrow{k_E^R} \mathrm{D^A} + \mathrm{D^A}, \quad ④ \ \mathrm{D^R} \xrightarrow{\delta + \bar{k}_E^R} \mathrm{D^A}, \end{aligned}$$

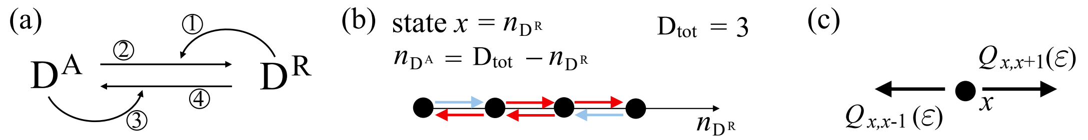

Figure 1: **1D model.** (a) Chemical reaction system. The numbers on the arrows correspond to the chemical reactions associated with the arrows as described in (2.1) in the main text. (b) Markov chain graph. Here, we consider $\mathrm{D_{tot}} = 3$ and we use black dots to represent the states, red arrows to represent transition rates that are $O(1)$, and blue arrows to represent transition rates that are $O(\varepsilon)$. (c) Directions of the potential transitions of X^ε starting from a state x, whose rates are given in equation (2.2).

where $\delta, k_E^A, \bar{k}_E^A, k_E^R, \bar{k}_E^R > 0$. Here, the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. We denote the reaction volume by V, and let $\varepsilon := \frac{\delta+\bar{k}_E^A}{k_E^A(\mathrm{D_{tot}}/V)} = \frac{\delta_A}{k_E^A(\mathrm{D_{tot}}/V)}$, where $\delta_A := \delta + \bar{k}_E^A$. We also consider the constant $\mu := \frac{k_E^R}{k_E^A}$, which captures the asymmetry between the erasure rates of repressive and activating histone modifications.We introduce the constant b such that $\mu b = \frac{\delta_R}{\delta_A}$, with $\delta_R := \delta + \bar{k}_E^R$. Then, $\delta_A = \varepsilon \frac{k_E^A \mathrm{D_{tot}}}{V}$ and $\delta_R := \delta_A \mu b = \varepsilon \frac{k_E^A \mathrm{D_{tot}}}{V} \mu b$. So, as $\varepsilon \to 0$, both δ_A and δ_R go to 0, with $\mathrm{D_{tot}}, \frac{k_E^A}{V}, \mu$, and b fixed.

Now, consider a continuous time Markov chain X^ε, with state space $\mathcal{X} := \{0, \ldots, \mathrm{D_{tot}}\}$, where $\mathrm{D_{tot}} \geq 2$ is an integer, which keeps track of $n_{\mathrm{D^R}}$ through time. Given that we have the conservation law $n_{\mathrm{D^R}} + n_{\mathrm{D^A}} = \mathrm{D_{tot}}$, $n_{\mathrm{D^A}}$ can be obtained as a function of $n_{\mathrm{D^R}}$, that is $n_{\mathrm{D^A}} = \mathrm{D_{tot}} - n_{\mathrm{D^R}}$. Assuming stochastic mass-action kinetics (including the usual volume scaling of rate constants [16]), the infinitesimal generator $Q(\varepsilon)$ [2] for X^ε is given by:

$$Q_{x,x+\ell}(\varepsilon) = \begin{cases} \left(\frac{k_E^A}{V} x + \varepsilon \frac{k_E^A}{V} \mathrm{D_{tot}}\right)(\mathrm{D_{tot}} - x) & \text{if } \ell = 1 \\ \mu \left(\frac{k_E^A}{V}(\mathrm{D_{tot}} - x) + b\varepsilon \frac{k_E^A}{V} \mathrm{D_{tot}}\right) x & \text{if } \ell = -1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

for $x \in \mathcal{X}$, $\ell \in \mathbb{Z} \setminus \{0\}$ and $x + \ell \in \mathcal{X}$, and $Q_{x,x}(\varepsilon) = -\sum_{y \in \mathcal{X} \setminus \{x\}} Q_{x,y}(\varepsilon)$ for $x \in \mathcal{X}$. We extend this definition to $\varepsilon = 0$ by defining $Q_{x,y}(0) := \lim_{\varepsilon \to 0} Q_{x,y}(\varepsilon)$ for $x, y \in \mathcal{X}$. We will follow a similar convention for other examples. We consider X^0 to be the continuous time Markov chain with infinitesimal generator given by $Q(0)$. The process X^0 corresponds to a SCRN model associated with the autocatalytic reactions ① and ③ in (2.1), alone. Note that

$$Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)}, \qquad \varepsilon \geq 0, \tag{2.3}$$

for appropriate matrices $Q^{(0)}$ and $Q^{(1)}$ in $\mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$. By (2.3), we can see that $Q(\cdot)$ is a *real-analytic* (and moreover *linear*) perturbation of $Q^{(0)}$ (see Section 1.5 for definitions). Note that

[2] Note that $Q(\varepsilon)$ is sometimes called an infinitesimal transition matrix. The entries $Q_{x,y}(\varepsilon)$ for $x \neq y$ are the infinitesimal transition rates of going from x to y: $\mathbb{P}[X^\varepsilon(t+h) = y | X^\varepsilon(t) = x] = Q_{x,y}(\varepsilon)h + o(h)$ as $h \to 0$.

for every $\varepsilon > 0$, X^ε is irreducible, while X^0 has a transient communicating class $\{1, \dots, \mathrm{D_{tot}} - 1\}$ and two absorbing states (0 and $\mathrm{D_{tot}}$) (see SI - Section S.8). Because of this discontinuity at $\varepsilon = 0$, we say that $Q(\cdot)$ is a *singular perturbation* of $Q^{(0)}$ (see Section 3.1 for a precise definition).

We first want to determine the probability for the gene to be in the active state a ($x = 0$), repressed state r ($x = \mathrm{D_{tot}}$) or one of the intermediate states ($x \in \{1, \dots, \mathrm{D_{tot}} - 1\}$) after a long time (life-time of the organism), as a function of ε. We are especially interested in the limit of the stationary distribution for the system, $\pi(\varepsilon)$, as $\varepsilon \to 0$ (i.e., the basal erasure rate of the chromatin modifications is much lower than their maintenance rate). Since X^ε is irreducible for $\varepsilon > 0$ (and it has a finite state space), it has a unique stationary distribution $\pi(\varepsilon)$. In Section 3.1 we show that $\pi(0) := \lim_{\varepsilon\to 0} \pi(\varepsilon)$ exists and the function $\pi(\cdot)$ admits a convergent power series expansion:

$$\pi(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)} \;\text{ for } 0 \le \varepsilon < \varepsilon_1, \tag{2.4}$$

for some $\varepsilon_1 > 0$. In order to determine $\pi(0)$, we can take limits and observe that $\pi(0)Q(0) = 0$ and so $\pi(0)$ is a stationary distribution for $Q(0)$. Indeed, $\pi(0)$ is a specific mixture of atoms on the two absorbing states (0 and $\mathrm{D_{tot}}$) for X^0.

In Figure 2 we see how the function $\pi(\varepsilon)$ changes as $\varepsilon \to 0$ for several values of μ with $\mathrm{D_{tot}}$, $\frac{k_E^A}{V}$ and b fixed. Furthermore, for this simpler chromatin modification circuit, because of the birth-death structure of X^ε, we can obtain explicit formulas for $\pi(\varepsilon)$ when $\varepsilon > 0$ (see SI - Section S.8). On letting $\varepsilon \to 0$, we obtain:

$$\pi_x(0) = \begin{cases} \frac{b\mu^{\mathrm{D_{tot}}}}{1+b\mu^{\mathrm{D_{tot}}}} & \text{if } x = 0 \\ 0 & \text{if } x \in \{1, \dots, \mathrm{D_{tot}} - 1\} \\ \frac{1}{1+b\mu^{\mathrm{D_{tot}}}} & \text{if } x = \mathrm{D_{tot}}. \end{cases} \tag{2.5}$$

Thus, $\pi_x(0) \neq 0$ only for $x = 0$ and $x = \mathrm{D_{tot}}$ and $\pi_0(0)$ increases as μ increases, while $\pi_{\mathrm{D_{tot}}}(0)$ decreases as μ increases.

For continuous time Markov chains beyond the one-dimensional birth-death processes seen here, determining $\pi(0)$ will be a considerable task. In Section 4.1, we address the problem of determining $\pi(0)$, together with the whole expansion (2.4), in a systematic way, for a class of singularly perturbed Markov chains that includes our models of chromatin modification circuits. For the 1D model considered here, the derivation of the first two terms in the expansion is given in Section 4.1.2.

Now, in order to evaluate the time to memory loss of the active and repressed states, let us define the first passage time as $\tau_y^\varepsilon = \inf\{t \ge 0 : \; X^\varepsilon(t) = y\}$ for a state $y \in \mathcal{X}$. We will see in (3.4) that the mean first passage time (MFPT) for X^ε starting from $x \in \mathcal{X}$, $h_{x,y}(\varepsilon) = \mathbb{E}_x[\tau_y^\varepsilon]$, has a Laurent series expansion of the form:

$$h_{x,y}(\varepsilon) = \frac{c_{-p}}{\varepsilon^p} + \dots + \frac{c_{-1}}{\varepsilon} + c_0 + \varepsilon c_1 + \dots \qquad \text{for } 0 < \varepsilon < \varepsilon_{\{y\}}, \tag{2.6}$$

for some $\varepsilon_{\{y\}} > 0$, for some natural number $p \ge 0$ and where $c_{-p} \neq 0$. Then, considering the repressed state $r = \mathrm{D_{tot}}$ and the active state $a = 0$, we define the time to memory loss of the

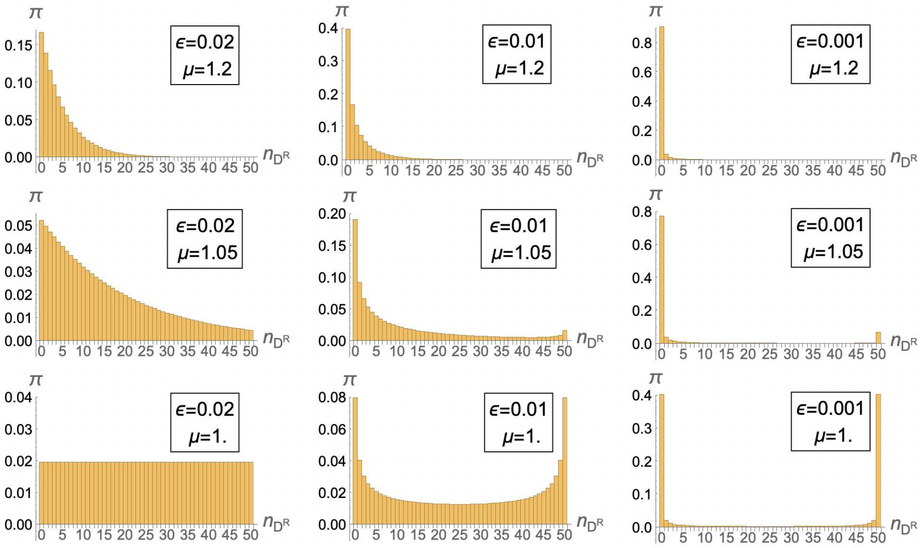

Figure 2: Histograms for the stationary distribution $\pi(\varepsilon)$ of the Markov chain X^ε defined by (2.2), for different values of ε and μ. The plot was generated by numerically solving $\pi(\varepsilon)Q(\varepsilon) = 0$ using the *Eigenvector* function in Mathematica. The parameters used were $\mathrm{D_{tot}} = 50, k_E^A/V = 1$, and $b = 1$.

repressed state as $h_{r,a}(\varepsilon)$ and the time to memory loss of the active state as $h_{a,r}(\varepsilon)$. Now, we are interested in the derivation of analytical formulas for $h_{r,a}(\varepsilon)$ and $h_{a,r}(\varepsilon)$. This will allow us to understand how the time to memory loss changes as $\varepsilon \to 0$, and how the asymmetry of the system, represented by μ, affects this limit. For this case study, exploiting its birth-death structure, we can directly derive relevant formulas (see SI - Section S.8, SI - Equations (S.65)-(S.66)). In particular, defining $\lambda_x^\varepsilon = Q_{x,x+1}(\varepsilon)$, $\gamma_x^\varepsilon = Q_{x,x-1}(\varepsilon)$, with $Q_{x,x+1}(\varepsilon)$ and $Q_{x,x-1}(\varepsilon)$ defined in (2.2), and $r_j^\varepsilon = \frac{\lambda_1^\varepsilon \lambda_2^\varepsilon \ldots \lambda_j^\varepsilon}{\gamma_1^\varepsilon \gamma_2^\varepsilon \ldots \gamma_j^\varepsilon}$, for $j = 1, 2, ..., \mathrm{D_{tot}} - 1$, the time to memory loss of the repressed state is given by

$$(2.7) \qquad h_{r,a}(\varepsilon) = \frac{r^\varepsilon_{\mathrm{D_{tot}}-1}}{\gamma^\varepsilon_{\mathrm{D_{tot}}}} \left(1 + \sum_{i=1}^{\mathrm{D_{tot}}-1} \frac{1}{r_i^\varepsilon}\right) + \sum_{i=2}^{\mathrm{D_{tot}}-1} \left[\frac{r^\varepsilon_{i-1}}{\gamma_i^\varepsilon} \left(1 + \sum_{j=1}^{i-1} \frac{1}{r_j^\varepsilon}\right)\right] + \frac{1}{\gamma_1^\varepsilon}.$$

Similarly, defining $\tilde{r}_j^\varepsilon = \frac{\gamma^\varepsilon_{\mathrm{D_{tot}}-1} \gamma^\varepsilon_{\mathrm{D_{tot}}-2} \ldots \gamma^\varepsilon_{\mathrm{D_{tot}}-j}}{\lambda^\varepsilon_{\mathrm{D_{tot}}-1} \lambda^\varepsilon_{\mathrm{D_{tot}}-2} \ldots \lambda^\varepsilon_{\mathrm{D_{tot}}-j}}$, for $j = 1, 2, ..., \mathrm{D_{tot}} - 1$, the time to memory loss of the active state is given by

$$(2.8) \qquad h_{a,r}(\varepsilon) = \frac{\tilde{r}^\varepsilon_{\mathrm{D_{tot}}-1}}{\lambda_0^\varepsilon} \left(1 + \sum_{j=1}^{\mathrm{D_{tot}}-1} \frac{1}{\tilde{r}_i^\varepsilon}\right) + \sum_{i=2}^{\mathrm{D_{tot}}-1} \left[\frac{\tilde{r}^\varepsilon_{i-1}}{\lambda^\varepsilon_{\mathrm{D_{tot}}-i}} \left(1 + \sum_{j=1}^{i-1} \frac{1}{\tilde{r}_j^\varepsilon}\right)\right] + \frac{1}{\lambda^\varepsilon_{\mathrm{D_{tot}}-1}}.$$

Since λ_0^ε and $\gamma^\varepsilon_{\mathrm{D_{tot}}}$ are the only transition rates that are $O(\varepsilon)$ with the rest being $O(1)$, the time to memory loss of both the active and repressed states are $O(\varepsilon^{-1})$, that is, $p = 1$, and as $\varepsilon \to 0$, these mean times tend to infinity.

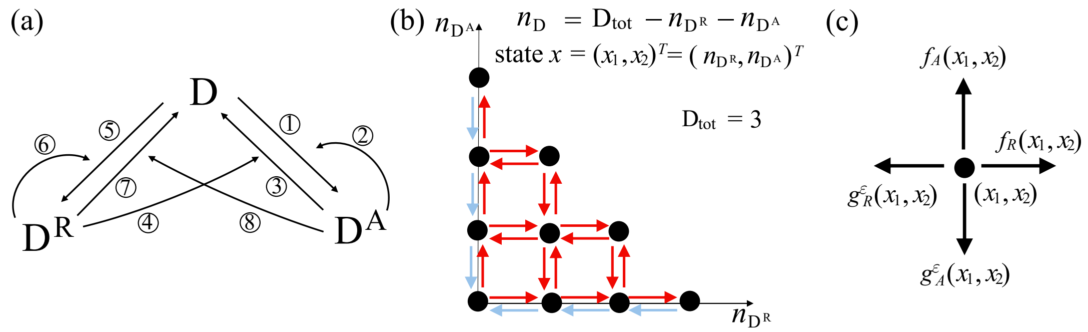

Figure 3: **2D model.** (a) Chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (2.9) in the main text. (b) Markov chain graph. Here, we consider $D_{tot} = 3$ and we use black dots to represent the states, red arrows to represent transition rates that are $O(1)$, and blue arrows to represent transition rates that are $O(\varepsilon)$. (c) Directions of the possible one step transitions for X^{ε} starting from a state $x = (x_1, x_2)^T$, whose rates are given in equation (2.10).

Furthermore, γ^{ε}_x, with $x \in \{1, 2, ..., D_{tot}\}$, are the only rates that depend on μ (they are linear in μ). Examining (2.7) and (2.8) with this observation in mind, we see that, if μ is increased (that is, the erasure rate of the repressive histone modification is increased compared to that of the active histone modification), $h_{a,r}(\varepsilon)$ increases, while $h_{r,a}(\varepsilon)$ decreases. The opposite happens when μ is decreased.

More complicated situations arise when we do not have a birth-death structure to work with, as in the model of the next example. To evaluate how critical system parameters affect the time to memory loss for such more elaborate systems, in Section 4, we develop an algorithm to determine p (see Section 4.2.1), we give an expression for the leading term in the series expansion of the mean first passage time, and we exploit theoretical results developed in our paper [13] for comparing continuous time Markov chains, to determine how the asymmetry of the system affects the time to memory loss (see Section 4.3).

2.2 2D model Let us consider a model in which, compared to the previous one, we assume that a nucleosome can also be unmodified. More precisely, in this case we denote the number of nucleosomes unmodified (D), modified with repressive modifications (D^R), and modified with activating modifications (D^A) by n_D, n_{D^R} and n_{D^A}, respectively, and we have that $n_D + n_{D^R} + n_{D^A} = D_{tot}$, with D_{tot} representing the total number of nucleosomes within the gene. Furthermore, each histone modification autocatalyzes its own production and promotes the erasure of the other one [10, 17]. The chemical reaction system is the following:

$$\text{(2.9)}\quad \begin{array}{llll} ① \ D \xrightarrow{k^A_{W0}+k^A_W} D^A, & ② \ D + D^A \xrightarrow{k^A_M} D^A + D^A, & ③ \ D^A \xrightarrow{\delta + \bar{k}^A_E} D, & ④ \ D^A + D^R \xrightarrow{k^A_E} D + D^R, \\ ⑤ \ D \xrightarrow{k^R_{W0}+k^R_W} D^R, & ⑥ \ D + D^R \xrightarrow{k^R_M} D^R + D^R, & ⑦ \ D^R \xrightarrow{\delta + \bar{k}^R_E} D, & ⑧ \ D^R + D^A \xrightarrow{k^R_E} D + D^A, \end{array}$$

where $k^A_{W0}, k^A_W, k^A_M, \delta, \bar{k}^A_E, k^A_E, k^R_{W0}, k^R_W, k^R_M, \bar{k}^R_E, k^R_E > 0$. Here, the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. Now, similarly to what we did for the previous model, let us denote the reaction volume by V, and let $\varepsilon := \frac{\delta+\bar{k}_E^A}{k_M^A(\mathrm{D_{tot}}/V)} = \frac{\delta_A}{k_M^A(\mathrm{D_{tot}}/V)}$, with $\delta_A := \delta + \bar{k}_E^A$, and $\mu := \frac{k_E^R}{k_E^A}$. Additionally, consider the constant b such that $\mu b = \frac{\delta_R}{\delta_A}$, with $\delta_R := \delta + \bar{k}_E^R$. Then $\delta_R = \delta_A \mu b = \varepsilon \frac{k_M^A \mathrm{D_{tot}}}{V} \mu b$. So, as $\varepsilon \to 0$, both δ_A and δ_R go to 0 with $\mathrm{D_{tot}}$, $\frac{k_M^A \mathrm{D_{tot}}}{V}$, μ and b fixed.

We consider the continuous time Markov chain $X^\varepsilon = \{(X_1^\varepsilon(t), X_2^\varepsilon(t))^T, t \geq 0\}$, which keeps track of $(n_{\mathrm{DR}}, n_{\mathrm{DA}})$ through time. Since the total number of nucleosomes $\mathrm{D_{tot}}$ is constant, the state space is $\mathcal{X} = \{x = (x_1, x_2)^T \in \mathbb{Z}_+^2 : x_1 + x_2 \leq \mathrm{D_{tot}}\}$. The potential one step transitions for X^ε from $x \in \mathcal{X}$ are shown in Figure 3(c), where the associated transition vectors are given by $v_1 = -v_2 = (0,1)^T$ and $v_3 = -v_4 = (1,0)^T$ and the infinitesimal transition rates (assuming mass-action kinetics with the usual volume scaling of rate constants) are given by

(2.10)
$$\begin{aligned} Q_{x,x+v_1}(\varepsilon) &= f_A(x) = (\mathrm{D_{tot}} - (x_1+x_2))\left(k_{W0}^A + k_W^A + \frac{k_M^A}{V}x_2\right), \\ Q_{x,x+v_3}(\varepsilon) &= f_R(x) = (\mathrm{D_{tot}} - (x_1+x_2))\left(k_{W0}^R + k_W^R + \frac{k_M^R}{V}x_1\right), \\ Q_{x,x+v_2}(\varepsilon) &= g_A^\varepsilon(x) = x_2\left(\varepsilon\frac{k_M^A}{V}\mathrm{D_{tot}} + x_1\frac{k_E^A}{V}\right), \quad Q_{x,x+v_4}(\varepsilon) = g_R^\varepsilon(x) = x_1\mu\left(\varepsilon\frac{k_M^A}{V}\mathrm{D_{tot}}b + x_2\frac{k_E^A}{V}\right). \end{aligned}$$

This is a more complicated model compared to the previous example and, in order to study its stationary distribution and mean first passage times, we will exploit the theory developed in this paper, as shown in Section 4.

3 Basic Setup and Definitions

In Section 3.1 we provide basic definitions for singularly perturbed continuous time Markov chains and describe some key properties for them. In particular, we describe the form of series expansions for their stationary distributions and mean first passage times. We will study these quantities and apply our results to a class of continuous time Markov chains called Stochastic Chemical Reaction Networks (SCRNs) which are defined in Section 3.2. Our models of chromatin modification circuits will be SCRNs. All of the models considered will have a finite state space.

3.1 Singularly perturbed, finite state, continuous time Markov chains

Suppose $\mathcal{X}$ is a finite set and $|\mathcal{X}| > 1$. For a value $\varepsilon_0 > 0$, consider a family $\{X^\varepsilon : 0 \leq \varepsilon < \varepsilon_0\}$ of continuous time Markov chains with state space $\mathcal{X}$ and infinitesimal generators $\{Q(\varepsilon) : 0 \leq \varepsilon < \varepsilon_0\}$ where $\varepsilon \mapsto Q(\varepsilon)$ is a real-analytic perturbation of $Q(0)$. Thus,

$$Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)} + \varepsilon^2 Q^{(2)} + \cdots, \tag{3.1}$$

where $\{Q^{(k)} : k \geq 0\}$ is a family of $|\mathcal{X}| \times |\mathcal{X}|$ real-valued matrices such that $\sum_{k=0}^\infty \varepsilon^k \|Q^{(k)}\| < \infty$ for every $0 \leq \varepsilon < \varepsilon_0$. Assume that the continuous time Markov chains X^ε are irreducible for $0 < \varepsilon < \varepsilon_0$. In this context, the perturbation is **singular** when X^0 has more than one recurrent class. This notion of singular will be the focus of our attention although some of our work applies for the regular (non-singular) case too. All of our chromatin modification circuit models have associated singular continuous time Markov chains, where the perturbation is **linear**, i.e., $Q^{(k)} = 0$ for every $k \geq 2$.

When $0<\varepsilon<\varepsilon_0$, there is an equivalent characterization of X^ε using holding times with exponential parameters $\{q_x(\varepsilon)\}_{x\in\mathcal{X}}$ and a transition probability matrix $P(\varepsilon)$ for the **embedded discrete time Markov chain**. Specifically, for each $x\in\mathcal{X}$, $q_x(\varepsilon)=-Q_{x,x}(\varepsilon)\neq 0$, since X^ε is irreducible, and for all $x,y\in\mathcal{X}$, $P_{x,x}(\varepsilon)=0$, $P_{x,y}(\varepsilon)=\frac{Q_{x,y}(\varepsilon)}{q_x(\varepsilon)}$, for $y\neq x$ in $\mathcal{X}$. Note that $Q(\varepsilon)=\mathrm{diag}(q(\varepsilon))(P(\varepsilon)-I)$. The matrix $P(\varepsilon)$ has a power series expansion in ε for sufficiently small $0\le\varepsilon<\varepsilon_P$ for some $\varepsilon_P>0$ (the justification is similar to that for (3.7) below).

The first quantities we are interested in studying are mean first passage times. Consider a nonempty set $\mathcal{B}\subseteq\mathcal{X}$ such that $\mathcal{B}\neq\mathcal{X}$ and let

$$\tau^\varepsilon_{\mathcal{B}}:=\inf\{t\ge 0: X^\varepsilon(t)\in\mathcal{B}\}.$$

We define the **mean first passage time (MFPT)** (for X^ε) from $x\in\mathcal{X}$ to $\mathcal{B}$ as

$$h_{x,\mathcal{B}}(\varepsilon)=\mathbb{E}[\tau^\varepsilon_{\mathcal{B}}\mid X^\varepsilon(0)=x].$$

If $\mathcal{B}=\{y\}$ for some $y\in\mathcal{X}$, we adopt the notation: $h_{x,y}(\varepsilon):=h_{x,\{y\}}(\varepsilon)$. Using first step analysis (see (3.1) in [26]), for $0<\varepsilon<\varepsilon_0$,

$$h_{x,\mathcal{B}}(\varepsilon)=\begin{cases}0 & \text{if } x\in\mathcal{B}\\ \frac{1}{q_x(\varepsilon)}+\sum_{y\in\mathcal{X}}P_{x,y}(\varepsilon)h_{y,\mathcal{B}}(\varepsilon) & \text{if } x\in\mathcal{B}^c.\end{cases} \tag{3.2}$$

Now, define $P^{\mathcal{B}^c}(\varepsilon)$ and $Q^{\mathcal{B}^c}(\varepsilon)$ as the matrices obtained by removing the columns and rows of $P(\varepsilon)$ and $Q(\varepsilon)$, respectively, corresponding to states in $\mathcal{B}$. Then, by noting that $I-P^{\mathcal{B}^c}(\varepsilon)$ is invertible (see SI - Lemma S.3) and that $Q^{\mathcal{B}^c}(\varepsilon)=-\,\mathrm{diag}((q_x(\varepsilon))_{x\in\mathcal{B}^c})(I-P^{\mathcal{B}^c}(\varepsilon))$ is invertible, from (3.2), we obtain

$$h_{\mathcal{B}}(\varepsilon)=-(Q^{\mathcal{B}^c}(\varepsilon))^{-1}\mathbb{1}, \tag{3.3}$$

where $h_{\mathcal{B}}(\varepsilon):=(h_{x,\mathcal{B}}(\varepsilon))_{x\in\mathcal{B}^c}$, I is the identity matrix of dimension $|\mathcal{B}^c|$, and $\mathbb{1}$ is the vector of all 1's, of size $|\mathcal{B}^c|$. Proposition 1.2, yields that there is $0<\varepsilon_{\mathcal{B}}<\varepsilon_0$ such that $-(Q^{\mathcal{B}^c}(\varepsilon))^{-1}$ can be expanded as a matrix-valued Laurent series as in (1.3) for $0<\varepsilon<\varepsilon_{\mathcal{B}}$, and then for each $x\in\mathcal{B}^c$,

$$\mathbb{E}_x[\tau^\varepsilon_{\mathcal{B}}]=h_{x,\mathcal{B}}(\varepsilon)=\sum_{k=-p(x)}^{\infty}\rho_x^{(k)}\varepsilon^k,\quad 0<\varepsilon<\varepsilon_{\mathcal{B}}, \tag{3.4}$$

where $p(x)\ge 0$ is an integer, $\rho_x^{(-p(x))}>0$, $\rho_x^{(k)}\in\mathbb{R}$ for $k>-p(x)$, and the convergence is absolute convergence for $0<\varepsilon<\varepsilon_{\mathcal{B}}$. The quantity $p(x)$ will be called the **order of the pole** of (3.4). In Section 4.2.1 we will show how to find $p(x)$ by using an algorithm that uses the order, with respect to ε, of the transitions of the Markov chain X^ε.

A second quantity of interest is the stationary distribution for X^ε. For $0<\varepsilon<\varepsilon_0$, since X^ε is assumed to be irreducible and has finite state space, there is a unique **stationary distribution** $\pi(\varepsilon)=(\pi_x(\varepsilon))_{x\in\mathcal{X}}$, which is the unique probability row vector satisfying $\pi(\varepsilon)Q(\varepsilon)=0$. We are interested in studying $\pi(\varepsilon)$ as $\varepsilon\to 0$. For this, first consider $\eta^\varepsilon_x=\inf\{t\ge 0: X^\varepsilon(t)\neq x\}$

and $\zeta_x^\varepsilon = \inf\{t > \eta_x^\varepsilon : X^\varepsilon(t) = x\}$, $x \in \mathcal{X}$. Note that $\mathbb{E}_y[\zeta_x^\varepsilon] = h_{y,x}(\varepsilon)$ for $y \neq x$. For each $x \in \mathcal{X}$, $\mathbb{E}_x[\zeta_x^\varepsilon]$ is called the **mean return time** to the state x, and for $0 < \varepsilon < \varepsilon_0$ satisfies

$$\mathbb{E}_x[\zeta_x^\varepsilon] = \frac{1}{q_x(\varepsilon)} + \sum_{y\neq x} P_{xy}(\varepsilon)\mathbb{E}_y[\zeta_x^\varepsilon] = \frac{1}{q_x(\varepsilon)} + \sum_{y\neq x} P_{xy}(\varepsilon) h_{y,x}(\varepsilon). \tag{3.5}$$

It is well known (see Theorem 3.8.1 in [26]) that for $0 < \varepsilon < \varepsilon_0$,

$$\pi_x(\varepsilon) = \frac{1}{\mathbb{E}_x[\zeta_x^\varepsilon]} \cdot \frac{1}{q_x(\varepsilon)}, \quad x \in \mathcal{X}. \tag{3.6}$$

From (3.4) and (3.5), we can see that $\varepsilon \mapsto q_x(\varepsilon)\mathbb{E}_x[\zeta_x^\varepsilon]$ can be extended to an analytic function on a punctured disk about 0 in $\mathbb{C}$, with a Laurent series expansion having at most a pole of finite order at 0. The radius of the punctured disk may be smaller than ε_0. This, together with (3.6), implies that $\varepsilon \mapsto \pi_x(\varepsilon)$ can be extended to an analytic function on a punctured disk about 0 in $\mathbb{C}$, also with a Laurent series expansion. Since this function is bounded by one when restricted to sufficiently small positive values of ε, we can remove the singularity at 0 and obtain that $\pi(0) := \lim_{\varepsilon\to 0} \pi(\varepsilon)$ exists and furthermore $\varepsilon \mapsto \pi(\varepsilon)$ is a real-analytic perturbation of $\pi(0)$. In other words,

$$\pi(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)}, \qquad 0 \leq \varepsilon < \varepsilon_1, \tag{3.7}$$

for sufficiently small $\varepsilon_1 > 0$ and where $\{\pi^{(k)} : k \geq 0\}$ is a sequence of real-valued $|\mathcal{X}|$-dimensional vectors such that $\sum_{k=0}^{\infty} \varepsilon^k \|\pi^{(k)}\| < \infty$ for every $0 \leq \varepsilon < \varepsilon_1$.

3.2 Stochastic Chemical Reaction Networks (SCRNs)

In this section, we provide some background on Stochastic Chemical Reaction Networks. The reader is referred to Anderson & Kurtz [4] for a more in depth introduction to this subject.

We assume there is a finite non-empty set $\mathscr{S} = \{\mathrm{S}_1, \ldots, \mathrm{S}_d\}$ of d **species**, and a finite non-empty set $\mathscr{R} \subseteq \mathbb{Z}_+^d \times \mathbb{Z}_+^d$ that represents chemical **reactions**. We assume that $(w, w) \notin \mathscr{R}$ for every $w \in \mathbb{Z}_+^d$. The set $\mathscr{S}$ represents d different molecular species in a system subject to reactions $\mathscr{R}$ which change the number of molecules of some species. For each $(v^-, v^+) \in \mathscr{R}$, the d-dimensional vector v^- (the **reactant vector**) counts how many molecules of each species are consumed in the reaction, while v^+ (the **product vector**) counts how many molecules of each species are produced. The reaction is usually written as

$$\sum_{i=1}^{d} (v^-)_i \mathrm{S}_i \longrightarrow \sum_{i=1}^{d} (v^+)_i \mathrm{S}_i. \tag{3.8}$$

To avoid the use of unnecessary species, we will assume that for each $1 \leq i \leq d$, there exists a vector $w = (w_1, \ldots, w_d)^T \in \mathbb{Z}_+^d$ with $w_i > 0$ such that (w, v) or (v, w) is in $\mathscr{R}$ for some $v \in \mathbb{Z}_+^d$, i.e., each species is either a reactant or a product in some reaction.

The net change in the quantity of molecules of each species due to a reaction $(v^-, v^+) \in \mathscr{R}$ is described by $v^+ - v^-$ and it is called the associated **reaction vector**. We denote the set of reaction vectors by $\mathcal{V} := \{v \in \mathbb{Z}^d \,|\, v = v^+ - v^- \text{ for some } (v^-, v^+) \in \mathscr{R}\}$, we let $n := |\mathcal{V}|$ the size of $\mathcal{V}$ and we enumerate the members of $\mathcal{V}$ as $\{v_1, \dots, v_n\}$. Note that $\mathcal{V}$ does not contain the zero vector because $\mathscr{R}$ has no elements of the form (w, w). Different reactions might have the same reaction vector. For each $v_j \in \mathcal{V}$ we consider the set $\mathscr{R}_{v_j} := \{(v^-, v^+) \in \mathscr{R} \mid v_j = v^+ - v^-\}$. The matrix $S \in \mathbb{R}^{d\times|\mathscr{R}|}$ whose columns are the elements $v^+ - v^-$ for $(v^-, v^+) \in \mathscr{R}$ will be called the **stoichiometric matrix**.

Given $(\mathscr{S}, \mathscr{R})$ we will consider an associated continuous time Markov chain $X = (X_1, \dots, X_d)^T$, with a state space $\mathcal{X}$ contained in $\mathbb{Z}^d_+$, which tracks the number of molecules of each species over time. Roughly speaking, the dynamics of X will be given by the following: given a current state $x = (x_1, \dots, x_d)^T \in \mathcal{X} \subseteq \mathbb{Z}^d_+$, for each reaction $(v^-, v^+) \in \mathscr{R}$, there is a clock which will ring at an exponentially distributed time (with rate $\Lambda_{(v^-,v^+)}(x)$). The clocks for distinct reactions are independent of one another. If the clock corresponding to $(v^-, v^+) \in \mathscr{R}$ rings first, the system moves from x to $x + v^+ - v^-$ at that time, and then the process repeats. We now define the continuous time Markov chain in more detail.

Consider sets of species $\mathscr{S}$ and reactions $\mathscr{R}$, a non-empty set $\mathcal{X} \subseteq \mathbb{Z}^d_+$ and a collection of functions $\Lambda = \{\Lambda_{(v^-,v^+)} : \mathcal{X} \longrightarrow \mathbb{R}_+\}_{(v^-,v^+)\in\mathscr{R}}$ such that for each $x \in \mathcal{X}$ and $(v^-, v^+) \in \mathscr{R}$, if $x + v^+ - v^- \notin \mathcal{X}$, then $\Lambda_{(v^-,v^+)}(x) = 0$. Now, for $1 \le j \le n$, define

$$\Upsilon_j(x) := \sum_{(v^-,v^+)\in\mathscr{R}_{v_j}} \Lambda_{(v^-,v^+)}(x). \tag{3.9}$$

Note that for each $x \in \mathcal{X}$ and $1 \le j \le n$, if $x + v_j \notin \mathcal{X}$, then $\Upsilon_j(x) = 0$. The functions $\{\Lambda_{(v^-,v^+)} : \mathcal{X} \longrightarrow \mathbb{R}_+\}_{(v^-,v^+)\in\mathscr{R}}$ are called **propensity** or **intensity** functions. A common form for the propensity functions is the following, which is associated with **mass action kinetics**:

$$\Lambda_{(v^-,v^+)}(x) = \kappa_{(v^-,v^+)} \prod_{i=1}^{d} (x_i)_{(v^-)_i}, \tag{3.10}$$

where $\{\kappa_{(v^-,v^+)}\}_{(v^-,v^+)\in\mathscr{R}}$ are non-negative constants and for $m, \ell \in \mathbb{Z}_+$, the quantity $(m)_\ell$ is the falling factorial, i.e., $(m)_0 := 1$ and $(m)_\ell := m(m-1)\dots(m-\ell+1)$.

A **stochastic chemical reaction network (SCRN)** (associated with $(\mathscr{S}, \mathscr{R}, \mathcal{X}, \Lambda)$) is a continuous time Markov chain X with state space $\mathcal{X}$ and infinitesimal generator Q given for $x, y \in \mathcal{X}$ by

$$Q_{x,y} = \begin{cases} \Upsilon_j(x) & \text{if } y - x = v_j \text{ for some } 1 \le j \le n, \\ -\sum_{j=1}^{n} \Upsilon_j(x) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

A SCRN associated with $(\mathscr{S}, \mathscr{R}, \mathcal{X}, \Lambda)$ is said to satisfy a **conservation law** if there is a d-dimensional non-zero vector m such that $m^T S = 0$, and hence $m^T X(t) = x_{\text{tot}}$ for every $t \ge 0$, for some constant x_{tot}. Consequently, we can reduce the dimension of the continuous time Markov chain describing the system by one. For example, if $m = (1, \dots, 1)^T$, then the projected process $(X_1, \dots, X_{d-1})^T$ is again a continuous time Markov chain with state space $\{(x_1, \dots, x_{d-1})^T \in \mathbb{Z}^{d-1}_+ \mid (x_1, \dots, x_{d-1}, x_{\text{tot}} - \sum_{i=1}^{d-1} x_i)^T \in \mathcal{X}\}$. In our examples, we will often use this type of reduction.

4 Main Results

In this section we describe the main theoretical results of this paper, under assumptions that go beyond those of our models of chromatin modification circuits. More precisely, we present results on stationary distributions and mean first passage times in Sections 4.1 and 4.2. Then, in Section 4.3 we exploit theoretical results developed in our companion work [13] to study monotonic dependence on parameters for a class of continuous time Markov chains related to chromatin modification circuits and other SCRNs.

4.1 Stationary distributions

This section focuses on characterizing the terms in the series expansion (3.7). In Section 4.1.1 we focus on determining the term $\pi^{(0)} = \pi(0)$, while in SI - Section S.2.1 we provide a result which enables computation of all of the higher order terms $\pi^{(k)}$, for $k > 0$, under additional assumptions. In Section 4.1.2 we apply these results to the examples introduced in Section 2. Additional characterizations of $\pi(0)$ and $\pi^{(1)}$ are given in the SI - Sections S.2.3 and S.2.4. Further examples for higher dimensional models of the chromatin modification circuits will be given in Section 5. We remind the reader that to ease notation, we have adopted the convention that stationary distribution vectors will be row vectors, even though we do not use the transpose notation T to indicate this.

4.1.1 The zeroth order term

As in Section 3.1, consider a family $\{X^\varepsilon : 0 \le \varepsilon < \varepsilon_0\}$ of continuous time Markov chains on a finite state space $\mathcal{X}$, with infinitesimal generators $\{Q(\varepsilon) : 0 \le \varepsilon < \varepsilon_0\}$ where $\varepsilon \mapsto Q(\varepsilon)$ is a real-analytic perturbation of $Q(0)$ with coefficients $\{Q^{(k)} : k \ge 0\}$ and additionally $Q(\varepsilon)$ is irreducible for every $0 < \varepsilon < \varepsilon_0$. The matrix $Q(0) = Q^{(0)}$ is a Q-matrix for which $\mathcal{X}$ decomposes into recurrent (or ergodic) states $\mathcal{A}$ and transient states $\mathcal{T}$. From now on, we assume the following.

Assumption 4.1. *The set $\mathcal{A}$ consists of $|\mathcal{A}| \ge 1$ absorbing states for $Q(0)$, while $\mathcal{T}$ consists of $|\mathcal{T}| \ge 1$ transient states for $Q(0)$.*

In other words, in the dynamics of $Q(0)$ there is at least one transient state, at least one recurrent state and all the recurrent states are absorbing. Now, we label the state space starting with the states in $\mathcal{A}$ and followed by the ones in $\mathcal{T}$. For every $k \ge 0$, we can write $Q^{(k)}$ as

$$Q^{(k)} = \left(\begin{array}{c|c} A_k & S_k \\ \hline R_k & T_k \end{array}\right), \tag{4.1}$$

where $A_k \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$, $S_k \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{T}|}$, $R_k \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{A}|}$ and $T_k \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{T}|}$. In a similar fashion, we can write

$$Q(\varepsilon) = \left(\begin{array}{c|c} A(\varepsilon) & S(\varepsilon) \\ \hline R(\varepsilon) & T(\varepsilon) \end{array}\right), \tag{4.2}$$

for $0 \le \varepsilon < \varepsilon_0$, where $A(\varepsilon) \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{A}|}$, $S(\varepsilon) \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{T}|}$, $R(\varepsilon) \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{A}|}$ and $T(\varepsilon) \in \mathbb{R}^{|\mathcal{T}|\times|\mathcal{T}|}$. From Assumption 4.1, we obtain that

$$Q^{(0)} = Q(0) = \left(\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array}\right), \tag{4.3}$$

where T_0 is an invertible matrix (see SI - Lemma S.5).

For each $0 < \varepsilon < \varepsilon_0$, we denote by $\pi(\varepsilon) = (\pi_x(\varepsilon))_{x \in \mathcal{X}}$ the stationary distribution for $Q(\varepsilon)$. In Section 3.1, we showed that the limit $\pi(0) := \lim_{\varepsilon \to 0} \pi(\varepsilon)$ exists and that $\varepsilon \mapsto \pi(\varepsilon)$ is a real-analytic perturbation of $\pi(0)$ with expansion given by (3.7) for $0 \le \varepsilon < \varepsilon_1$. For convenience, decompose the row vector $\pi(\varepsilon)$ as $\pi(\varepsilon) = [\alpha(\varepsilon), \beta(\varepsilon)]$ for $0 \le \varepsilon < \varepsilon_1$ where $\alpha(\varepsilon) \in \mathbb{R}^{|\mathcal{A}|}$ and $\beta(\varepsilon) \in \mathbb{R}^{|\mathcal{T}|}$. From (3.7), letting $\pi^{(k)} = [\alpha^{(k)}, \beta^{(k)}]$, we have

$$\alpha(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \alpha^{(k)} \quad \text{and} \quad \beta(\varepsilon) = \sum_{k=0}^{\infty} \varepsilon^k \beta^{(k)}$$

for $0 \le \varepsilon < \varepsilon_1$. Since $\pi(\varepsilon)$ is a probability distribution for every $0 \le \varepsilon < \varepsilon_1$, we have that $\sum_{k=0}^{\infty} \varepsilon^k (\pi^{(k)} \mathbb{1}) = 1$, which yields that $\pi(0)\mathbb{1} = 1$ and $\pi^{(k)}\mathbb{1} = 0$ for every $k \ge 1$. Since $\pi(0)Q(0) = 0$, $\pi(0)$ is a stationary distribution for X^0 and so, by Assumption 4.1, it must be supported on $\mathcal{A}$ and so $\beta^{(0)} = 0$. In the next result we establish an equation that is satisfied by $\alpha^{(0)} = \alpha(0)$ and introduce a key matrix for our analysis. For convenience, let $\alpha := \alpha(0)$.

Lemma 4.1. *Under Assumption 4.1, $\pi(0) = [\alpha, 0]$, where 0 is the zero row vector of size $|\mathcal{T}|$ and α is an $|\mathcal{A}|$-dimensional probability vector satisfying the equation:*

$$\alpha(A_1 + S_1(-T_0)^{-1} R_0) = 0. \tag{4.4}$$

In addition,

$$\beta^{(1)} = \alpha S_1 (-T_0)^{-1}. \tag{4.5}$$

See SI - Section S.2.2 for the proof of Lemma 4.1. For convenience, we adopt the notation:

$$Q_{\mathcal{A}} := A_1 + S_1(-T_0)^{-1} R_0. \tag{4.6}$$

In SI - Lemma S.15, we show that $Q_{\mathcal{A}}$ is a Q-matrix of size $|\mathcal{A}| \times |\mathcal{A}|$. As a consequence, there exists a continuous time Markov chain with state space $\mathcal{A}$ and infinitesimal generator $Q_{\mathcal{A}}$. In general, a probability vector satisfying (4.4) needs not be unique. The following condition will imply uniqueness.

Assumption 4.2. *The Markov chain associated with $Q_{\mathcal{A}}$ has a single recurrent class.*

By SI - Lemma S.1, Assumption 4.2 is equivalent to the condition $\dim(\ker(Q_{\mathcal{A}}^T)) = 1$. The next result then follows from Lemma 4.1.

Theorem 4.2. *Suppose Assumptions 4.1 and 4.2 hold. Then, $\pi(0) = [\alpha, 0]$, where α is the unique probability vector on $\mathcal{A}$ such that $\alpha Q_{\mathcal{A}} = 0$.*

As we will see, all of the chromatin modification circuit models presented in this work satisfy both Assumptions 4.1 and 4.2. Also note that Lemma 4.1 yields a characterization of $\beta^{(1)}$ by means of (4.5).

Theorem 4.2 is simple to state, yet less easy to use since simple formulas for $Q_{\mathcal{A}}$ can be seldom obtained, making Assumption 4.2 hard to verify directly using (4.6). In this regard, we now introduce an auxiliary continuous time Markov chain $\tilde{X}$ and use it to construct (via time-change) a realization $\hat{X}_{\mathcal{A}}$ of the continuous time Markov chain with infinitesimal generator $Q_{\mathcal{A}}$. This will enable us to give assumptions on $\tilde{X}$ that will imply Assumption 4.2 and which can sometimes be easier to verify. Also, this explicit realization for $\hat{X}_{\mathcal{A}}$ can lead to alternative ways to verify Assumption 4.2. Under Assumption 4.1, consider the matrix

$$\tilde{Q} := \left(\begin{array}{c|c} A_1 & S_1 \\ \hline R_0 & T_0 \end{array}\right). \tag{4.7}$$

In SI - Lemma S.15 we prove that $\tilde{Q}$ is a Q-matrix. Let $\tilde{X}$ be a continuous time Markov chain with infinitesimal generator $\tilde{Q}$. For the purpose of illustration, if we assume that the perturbation is linear (as in (2.3)) and $\varepsilon_0 > 1$, then the transitions of $\tilde{X}$ consist of the transitions of X^0 augmented by the transitions of X^1 that emanate from $\mathcal{A}$. See Figure 4(a)-(b) for an illustration related to the 1D and 2D models, respectively, introduced in Section 2.

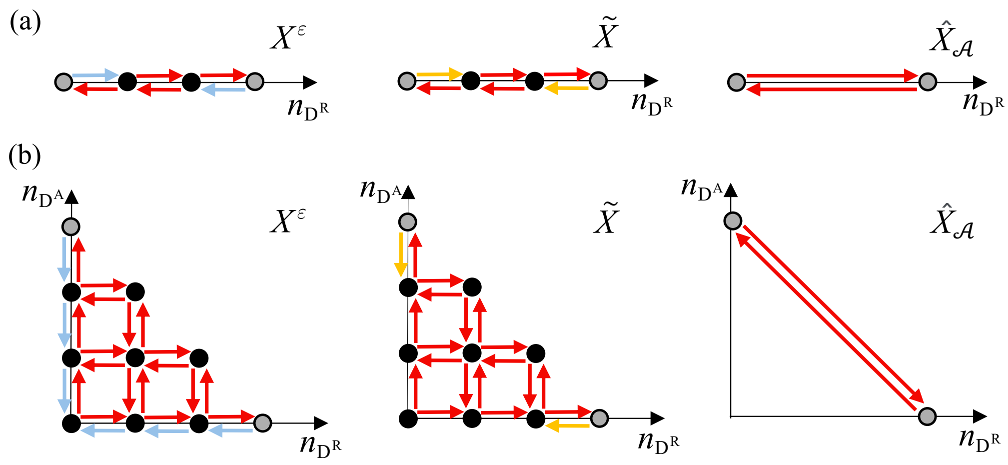

Figure 4: **Graphs for the one-step transitions of X^ε, $\tilde{X}$ and $\hat{X}_{\mathcal{A}}$ for the (a) 1D model and (b) 2D model.** Here, we consider $D_{tot} = 3$ and we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states, red arrows to represent transitions that are $O(1)$ for X^ε, $\tilde{X}$ and $\hat{X}_{\mathcal{A}}$, blue arrows to represent transitions that are $O(\varepsilon)$ for X^ε, and golden arrows to represent the transitions for $\tilde{X}$ that were $O(\varepsilon)$ for X^ε and became $O(1)$ for $\tilde{X}$.

Now, consider the **occupation time** of $\mathcal{A}$ by the Markov chain $\tilde{X}$ up to time $t \geq 0$, given by $\chi_{\mathcal{A}}(t) := \int_0^t \mathbb{1}_{\mathcal{A}}(\tilde{X}(s))ds$ for $t \geq 0$. Denote by $\chi_{\mathcal{A}}(\infty) = \lim_{t\to\infty} \chi_{\mathcal{A}}(t) = \int_0^\infty \mathbb{1}_{\mathcal{A}}(\tilde{X}(s))ds$. Since T_0 is invertible, SI - Lemmas S.5 and S.6 yield that $\mathbb{P}_x[\chi_{\mathcal{A}}(\infty) = \infty] = 1$ for all $x \in \mathcal{X}$. Additionally, consider the right-continuous inverse of $\chi_{\mathcal{A}}$, $\tau(s) := \inf\{t \geq 0 : \chi_{\mathcal{A}}(t) > s\}$, defined for $s \geq 0$. We define the **restriction** process $\hat{X}_{\mathcal{A}}$ as

$$\hat{X}_{\mathcal{A}}(s) := \tilde{X}(\tau(s)), \quad s \geq 0. \tag{4.8}$$

By properties of the right-continuous inverse (see Problem 4.5 in [23], for example), the reader may verify that $\hat{X}_{\mathcal{A}}$ corresponds to observing $\tilde{X}$ only on the time intervals where $\tilde{X}$ is in $\mathcal{A}$. Roughly speaking, we are *erasing* the times where $\tilde{X}$ is outside of $\mathcal{A}$. In the language of Blumenthal & Getoor [9], $\chi_{\mathcal{A}}$ is a continuous additive functional for $\tilde{X}$, and by Exercise V.2.11 in [9], we obtain that $\hat{X}_{\mathcal{A}}$ is a continuous time Markov chain with state space $\mathcal{A}$. In the next result, we prove that $\hat{X}_{\mathcal{A}}$ is a realization of the continuous time Markov chain associated with $Q_{\mathcal{A}}$. See Figure 4(a)-(b) for a representation of $\hat{X}_{\mathcal{A}}$ associated with the 1D and 2D models, respectively.

Lemma 4.3. *Suppose Assumption 4.1 holds. Then, $\hat{X}_{\mathcal{A}}$ has infinitesimal generator $Q_{\mathcal{A}}$.*

The proof of Lemma 4.3 is given in SI - Section S.2.2. We now introduce some assumptions that imply that Assumption 4.2 holds. In addition, these assumptions will allow for some refinements (see SI - Section S.2).

Assumption 4.3. *For $\tilde{X}$, there exists a communicating class $\mathcal{C}$ such that $\mathcal{A} \subseteq \mathcal{C}$.*

We note that, if such a class $\mathcal{C}$ exists, then it has to be recurrent. In fact, if it was transient then $\chi_{\mathcal{A}}(\infty) < \infty$ with positive probability under $\mathbb{P}_x$, $x \in \mathcal{A}$, which is a contradiction.

Assumption 4.4. *The Markov chain $\tilde{X}$ is irreducible.*

We note that Assumption 4.4 implies Assumption 4.3. Moreover, they are both related to Assumption 4.2 in the following way.

Lemma 4.4. *Suppose Assumptions 4.1 and 4.3 hold. Then, the process $\hat{X}_{\mathcal{A}}$ is irreducible. As a consequence, either of Assumptions 4.4 or 4.3 implies that Assumption 4.2 holds.*

The proof of Lemma 4.4 is given in SI - Section S.2.2. The next result follows from Lemmas 4.3, 4.4 and Theorem 4.2.

Theorem 4.5. *Suppose Assumptions 4.1 and 4.3 hold. Then, $\pi(0) = [\alpha, 0]$ where α is the unique stationary distribution for the process $\hat{X}_{\mathcal{A}}$ and all entries of α are strictly positive.*

Assumptions 4.3 and 4.4 can be understood graphically in some cases. For example, Figure 4 illustrates that for the 1D-model, Assumption 4.4 is satisfied. For the 2D-model, we can see that while Assumption 4.4 is not satisfied (since the state $(0, 0)$ forms its own (transient) class for $\tilde{X}$), Assumption 4.3 does indeed hold. In Section 5 we will see that neither Assumption 4.4 nor 4.3 is satisfied by the 3D or 4D model. However, the weaker Assumption 4.2 does hold.

In the SI, we give recursive formulae for the higher order terms $\pi^{(k)}$, $k = 1, 2, \ldots$, under the following additional assumption (see SI - Theorem S.9).

Assumption 4.5. *The perturbation is linear, i.e., $Q(\varepsilon) = Q^{(0)} + \varepsilon Q^{(1)}$ for $0 \leq \varepsilon < \varepsilon_0$.*

4.1.2 Illustrative examples: 1D and 2D model

1D model. We use the tools developed in the preceding section to derive the terms $\pi^{(0)}$ and $\pi^{(1)}$ in the expansion (2.4) for the 1D model introduced in Section 2.1. Fix $\mathrm{D}_{\mathrm{tot}} \geq 2$ and let X^{ε} with infinitesimal generator $Q(\varepsilon)$ be as in Section 2.1, with the expression for $Q^{(k)}$ given in (4.1). By (2.3), Assumption 4.5 holds. Moreover, for each $0 < \varepsilon < \varepsilon_0$, with ε_0 being a fixed, positive constant, $Q(\varepsilon)$ is irreducible, while $Q(0)$ has a non-empty set of transient states $\mathcal{T} = \{1, \ldots, \mathrm{D}_{\mathrm{tot}} - 1\}$ and a set of two absorbing states $\mathcal{A} = \{a, r\}$, with $a = 0$ representing the fully active state ($n_{\mathrm{D}^{\mathrm{A}}} = \mathrm{D}_{\mathrm{tot}}$) and $r = \mathrm{D}_{\mathrm{tot}}$ representing the fully repressed state ($n_{\mathrm{D}^{\mathrm{R}}} = \mathrm{D}_{\mathrm{tot}}$). Then, Assumption 4.1 holds (see SI - Section S.8). Furthermore, by defining $f(x) := x(\mathrm{D}_{\mathrm{tot}} - x)$ for $x \in \mathcal{X}$, we can write the matrices R_0 and T_0 in the matrix $Q^{(0)}$ as follows:

$$R_0 = \begin{pmatrix} \mu\frac{k_E^A}{V}f(1) & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \frac{k_E^A}{V}f(1) \end{pmatrix}, \quad T_0 = \begin{pmatrix} -(\lambda_1^0+\gamma_1^0) & \lambda_1^0 & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \vdots & \vdots \\ \dots & 0 & \gamma_x^0 & -(\lambda_x^0+\gamma_x^0) & \lambda_x^0 & 0 & \dots \\ \vdots & 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & \gamma^0_{\mathrm{D_{tot}}-1} & -(\lambda^0_{\mathrm{D_{tot}}-1}+\gamma^0_{\mathrm{D_{tot}}-1}) \end{pmatrix}$$

where R_0 is a $(\mathrm{D_{tot}}-1)\times 2$ matrix and T_0 is a $(\mathrm{D_{tot}}-1)\times(\mathrm{D_{tot}}-1)$ tridiagonal matrix, and $\gamma_x^0 = \mu\frac{k_E^A}{V}f(x)$, $\lambda_x^0 = \frac{k_E^A}{V}f(x)$, and $f(\mathrm{D_{tot}}-1) = f(1) = (\mathrm{D_{tot}}-1)$. In addition, we can write A_1 and S_1 of $Q^{(1)}$ as follows:

$$A_1 = \begin{pmatrix} -\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 & 0 \\ 0 & -b\mu\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 \end{pmatrix}, \quad S_1 = \begin{pmatrix} \frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & b\mu\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 \end{pmatrix}.$$

The process $\tilde{X}$, whose infinitesimal generator is defined in (4.7), is irreducible (see SI - Section S.8). This is illustrated in Figure 4(a). Thus, Assumption 4.4 holds. Then, Assumption 4.3 is also satisfied and Theorem 4.5 can be applied. This yields that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0\dots, 0]$ where α is the unique stationary distribution for the restriction process $\hat{X}_{\mathcal{A}}$ (defined by (4.8)), whose infinitesimal generator is $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$ by Lemma 4.3 and (4.6). Now,

(4.9)
$$Q_{\mathcal{A}} = \begin{pmatrix} -\frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 & \frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 \\ b\mu^{\mathrm{D_{tot}}}\frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 & -b\mu^{\mathrm{D_{tot}}}\frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 \end{pmatrix} = \frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2\begin{pmatrix} -1 & 1 \\ b\mu^{\mathrm{D_{tot}}} & -b\mu^{\mathrm{D_{tot}}} \end{pmatrix},$$

and since α is the unique probability vector satisfying $\alpha Q_{\mathcal{A}} = 0$, we have

$$\alpha_a = \frac{b\mu^{\mathrm{D_{tot}}}}{1+b\mu^{\mathrm{D_{tot}}}}, \quad \alpha_r = \frac{1}{1+b\mu^{\mathrm{D_{tot}}}}.$$

These results are in agreement with (2.5) in Section 2.1, where we explicitly computed the stationary distribution $\pi(\varepsilon)$ and let $\varepsilon \to 0$ (see SI - Section S.8).

Now, since Assumptions 4.1, 4.3, and 4.5 hold, we can apply SI - Theorem S.9 to derive an expression for $\beta^{(1)}$. For the transient states $\mathcal{T} = \{1, ..., \mathrm{D_{tot}}-1\}$, we have $\beta^{(1)} = [\pi_1^{(1)}, ..., \pi^{(1)}_{\mathrm{D_{tot}}-1}] = \alpha S_1(-T_0)^{-1}$, and so for $x \in \mathcal{T}$

$$\beta_x^{(1)} = \frac{b\mu^{\mathrm{D_{tot}}}}{1+b\mu^{\mathrm{D_{tot}}}}\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2(-T_0)^{-1}_{1,x} + \frac{1}{1+b\mu^{\mathrm{D_{tot}}}}b\mu\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2(-T_0)^{-1}_{\mathrm{D_{tot}}-1,x}, \tag{4.10}$$

in which $(-T_0)^{-1}_{1,x}$ and $(-T_0)^{-1}_{\mathrm{D_{tot}}-1,x}$, for $x \in \mathcal{T}$, are the elements indexed by $(1, x)$ and $(\mathrm{D_{tot}} - 1, x)$ of the matrix $(-T_0)^{-1}$, respectively. After some calculations, we obtain

$$(-T_0)^{-1}_{1,x} = \frac{\left(\frac{k_E^A}{V}\right)^{\mathrm{D_{tot}}-2} \frac{\prod_{i=1}^{\mathrm{D_{tot}}-1} f(i)}{f(x)} \left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1-x} \mu^i\right)}{\left(\frac{k_E^A}{V}\right)^{\mathrm{D_{tot}}-1} \prod_{i=1}^{\mathrm{D_{tot}}-1} f(i) \left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1} \mu^i\right)} = \frac{\left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1-x} \mu^i\right)}{\frac{k_E^A}{V} f(x) \left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1} \mu^i\right)},$$
$$(-T_0)^{-1}_{\mathrm{D_{tot}}-1,x} = \frac{\left(\frac{k_E^A}{V}\right)^{\mathrm{D_{tot}}-2} \frac{\prod_{i=1}^{\mathrm{D_{tot}}-1} f(i)}{f(x)} \left(1+\prod_{i=1}^{x-1} \mu^i\right)\mu^{\mathrm{D_{tot}}-1-x}}{\left(\frac{k_E^A}{V}\right)^{\mathrm{D_{tot}}-1} \prod_{i=1}^{\mathrm{D_{tot}}-1} f(i) \left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1} \mu^i\right)} = \frac{\left(1+\prod_{i=1}^{x-1} \mu^i\right)\mu^{\mathrm{D_{tot}}-1-x}}{\frac{k_E^A}{V} f(x) \left(1+\prod_{i=1}^{\mathrm{D_{tot}}-1} \mu^i\right)},$$

and then $\beta_x^{(1)}$, $x \in \mathcal{T}$, can be written as follows:

(4.11)
$$\beta_x^{(1)} = \frac{\mathrm{D}^2_{\mathrm{tot}}}{f(x)} \frac{b\mu^{\mathrm{D_{tot}}-x}}{1+b\mu^{\mathrm{D_{tot}}}} = \frac{\mathrm{D}^2_{\mathrm{tot}}}{x(\mathrm{D_{tot}}-x)} \frac{b\mu^{\mathrm{D_{tot}}-x}}{1+b\mu^{\mathrm{D_{tot}}}}.$$

2D model. In this section we analyze the stationary distribution for the 2D model introduced in Section 2.2. Fix $\mathrm{D_{tot}} \geq 2$ and let X^ε with infinitesimal generator $Q(\varepsilon)$ be as in Section 2.1, with the expression for the $Q^{(k)}$ given by (4.1). By (2.10), for this model Assumption 4.5 holds. Furthermore, $Q(0)$ has a non-empty set of transient states $\mathcal{T} = \{i_1, \dots, i_m\}$ where $m = \frac{(\mathrm{D_{tot}}+2)(\mathrm{D_{tot}}+1)}{2} - 2$, $i_1 = (0, \mathrm{D_{tot}}-1)^T$, $i_m = (\mathrm{D_{tot}}-1, 0)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, \mathrm{D_{tot}})^T$ corresponding to the fully active state ($n_{\mathrm{D^A}} = \mathrm{D_{tot}}$) and with $r = (\mathrm{D_{tot}}, 0)^T$ corresponding to the fully repressed state ($n_{\mathrm{D^R}} = \mathrm{D_{tot}}$), respectively. Then, Assumption 4.1 holds (see SI - Section S.9).

From (2.10), we see that $A_0 = 0$, $S_0 = 0$ and

$$A_1 = \begin{pmatrix} -\frac{k_M^A}{V}\mathrm{D}^2_{\mathrm{tot}} & 0 \\ 0 & -\frac{k_M^A}{V}\mathrm{D}^2_{\mathrm{tot}}\mu b \end{pmatrix}, \; S_1 = \begin{pmatrix} \frac{k_M^A}{V}\mathrm{D}^2_{\mathrm{tot}} & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & \frac{k_M^A}{V}\mathrm{D}^2_{\mathrm{tot}}\mu b \end{pmatrix}.$$

Furthermore, $R_0 \in \mathbb{R}^{m\times 2}$ is given by

$$\begin{pmatrix} f_A(0,\mathrm{D_{tot}}-1) & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & f_R(\mathrm{D_{tot}}-1,0) \end{pmatrix} = \begin{pmatrix} k_{W0}^A + k_W^A + \frac{k_M^A}{V}(\mathrm{D_{tot}}-1) & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0}^R + k_W^R + \frac{k_M^R}{V}(\mathrm{D_{tot}}-1) \end{pmatrix},$$

and $R_1 = 0$. The matrices T_0 and T_1 are more complex and examples of them, for $\mathrm{D_{tot}} = 2$, are provided in SI - Section S.9. For $\tilde{X}$, $\mathcal{C} = \mathcal{X} \setminus \{(0,0)\}$ is a communicating class such that $\mathcal{A} \subseteq \mathcal{C}$. This implies that Assumption 4.3 is satisfied. Given that Assumptions 4.1 and 4.3 are satisfied, Theorem 4.5 can be applied and we obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0 \dots, 0]$ where α is the unique stationary distribution for the process $\hat{X}_{\mathcal{A}}$, whose infinitesimal generator is $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$. This means that α is the unique probability vector such that $\alpha(A_1 + S_1(-T_0)^{-1}R_0) = 0$. Furthermore, given that Assumption 4.5 is satisfied, we can apply SI - Theorem S.9 to derive an expression for $\beta^{(1)} = [\pi^{(1)}_{i_1}, ..., \pi^{(1)}_{i_m}] = \alpha S_1(-T_0)^{-1}$. For example, if $\mathrm{D_{tot}} = 2$, the matrix $Q_{\mathcal{A}}$ is given by

$$Q_{\mathcal{A}} = \frac{4}{K}\frac{k_M^A}{V}\begin{pmatrix} -(k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) & (k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) \\ b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) & -b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}) \end{pmatrix}, \tag{4.12}$$

with

$$K = (k_{W0}^A + k_W^A + \frac{k_M^A}{V} + k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V}) + \mu(k_{W0}^R + k_W^R + \frac{k_M^R}{V} + k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V}), \tag{4.13}$$

and then $\pi^{(0)}$ is given by

$$\pi_x^{(0)} = \begin{cases} \frac{b\mu^2(k_{W0}^A+k_W^A)(k_{W0}^A+k_W^A+\frac{k_M^A}{V})}{b\mu^2(k_{W0}^A+k_W^A)(k_{W0}^A+k_W^A+\frac{k_M^A}{V})+(k_{W0}^R+k_W^R)(k_{W0}^R+k_W^R+\frac{k_M^R}{V})} & \text{if } x = (0, \mathrm{D_{tot}})^T \\ 0 & \text{if } x \in \mathcal{T} \\ \frac{(k_{W0}^R+k_W^R)(k_{W0}^R+k_W^R+\frac{k_M^R}{V})}{b\mu^2(k_{W0}^A+k_W^A)(k_{W0}^A+k_W^A+\frac{k_M^A}{V})+(k_{W0}^R+k_W^R)(k_{W0}^R+k_W^R+\frac{k_M^R}{V})} & \text{if } x = (\mathrm{D_{tot}}, 0)^T. \end{cases} \tag{4.14}$$

See SI - Section S.9 for the evaluation of $\pi_x^{(1)}$ for the transient states $x \in \mathcal{T}$ when $\mathrm{D_{tot}} = 2$. For this value of $\mathrm{D_{tot}}$, we see from (4.14) that $\pi_x^{(0)}$ depends monotonically on μ for each fixed x. As $\mathrm{D_{tot}}$ increases, the algebraic complexity of a full parameter representation of $\pi_x^{(0)}$ increases very rapidly. Thus, to investigate monotonic dependence on parameters for biologically relevant values of $\mathrm{D_{tot}}$ (of the order of 50, considering an average gene length of 10,000 bp [15] and one nucleosome per 200 bp [18]), we shall use comparison theorems developed in [13], without calculating any explicit formula (Section 4.3).

4.2 Mean first passage times (MFPTs)

In this section we develop a theoretical framework to study mean first passage times for continuous time Markov chains. We first develop an algorithm to determine the order of the pole of MFPTs for singularly perturbed Markov chains (Section 4.2.1). In Section 4.2.2, we focus on determining the leading coefficient for MFPTs, under some assumptions introduced in Section 4.1. In Section 4.2.3, we apply these results to the examples introduced in Section 2.

4.2.1 Algorithm to find the order of the poles for MFPTs

Our algorithm is adapted from an algorithm developed by Hassin and Haviv [20] for discrete time Markov chains. The idea used in [20] was to consider transitions between subsets of states and to keep track of the sojourn times in the sets of states. This is used to define a coarser version of the process, which may not be a Markov process and which moves between groups of states of the original Markov chain. This idea can be adapted to the continuous time setting as well. For this, we introduce stopping times to more explicitly track the sojourn times than was done in [20]. In addition, we extend the original algorithm's scope to consider the mean first passage time to a subset of states, instead of just a single state. The paper [20] uses r-cycles and notes that these could be replaced by more general r-components. Here, we focus on using the latter and call the set of vertices in such an r-component an r-connected set.

In this section, we consider a singularly perturbed, finite-state, continuous time Markov chain X^ε on $\mathcal{X}$ with infinitesimal generator $Q(\varepsilon)$ as described in Section 3.1. We provide an algorithm for finding the orders $\{p(v) : v \in \mathcal{B}^c\}$ of the poles for the mean first passage times to $\mathcal{B} \subset \mathcal{X}$ for X^ε starting from states in $\mathcal{B}^c$, where $\mathcal{B} \neq \emptyset$ is a strict subset of $\mathcal{X}$. We begin with a few definitions and some notation and then present the algorithm.

Definition 4.6. *Given $\varepsilon_0 > 0$ and a function $f : (0, \varepsilon_0) \to \mathbb{R}_{>0}$, we say $f = \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and strictly positive $m, M \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,*

$$m\varepsilon^k \leq f(\varepsilon) \leq M\varepsilon^k.$$

*If $f = \Theta(\varepsilon^k)$ for some $k \in \mathbb{Z}$, we say the **order** (at the origin) of f is k. If $f = \Theta(\varepsilon^{-k})$ where $k \in \mathbb{Z}_+$, we say that the **order of the pole** of f is k.*

Because the perturbation of X^ε is real analytic, $|\mathcal{X}| > 1$ and X^ε is irreducible for $\varepsilon > 0$, there exists $\varepsilon_{\max} > 0$ such that for each $x \neq y \in \mathcal{X}$, either $Q_{x,y}(\varepsilon) \equiv 0$ for all $\varepsilon \in (0, \varepsilon_{\max})$ or $Q_{x,y}(\varepsilon) > 0$ for all $\varepsilon \in (0, \varepsilon_{\max})$. In the latter case, the order of $Q_{x,y}(\varepsilon)$ is a non-negative integer, which we denote by k_{xy}. We let $E_0 = \{(x, y) : Q_{x,y}(\varepsilon) > 0 \text{ for all } \varepsilon \in (0, \varepsilon_{\max})\}$. As the algorithm progresses, states of $\mathcal{X}$ are gathered together to form composite nodes and the graph of the states of X^ε progresses through a series of reduced graphs. If u is a node in one of the graphs, then $S(u) \subset \mathcal{X}$ consists of the states in $\mathcal{X}$ that are collapsed to form the (reduced) node u. In Steps 2 and 3 of the algorithm, the function $\mathcal{K}$ and the initial values of p are inductively determined for all of these graphs. The final values of p for nodes in $\mathcal{B}^c$ are then determined in Step 4. With $\mathcal{K}_{uv}$ being defined, a directed edge (u, v) in one of the graphs is called an **r-edge**, where r is for regular, if $\mathcal{K}_{uv} = 0$, and an **r-path** is a directed path in the graph consisting of r-edges only. A set C in one of the graphs is called an **r-connected set** if $|C| > 1$ and there exists an r-path from u to v for any $u \neq v \in C$. The **order of the pole of the expected sojourn time** spent in an r-connected set C depends only on the set C and is denoted by $p(c)$ where c is a node representing the set C. For any node w outside of C, $\mathcal{K}_{cw}$ and $\mathcal{K}_{wc}$ are the **the order of the probabilities of a one-step transition** from c to w and from w to c, respectively. In Step 4 of the algorithm, $p(\cdot)$ keeps being updated but will stay finite and eventually fixate. The algorithm statement and related proof can be found in the SI - Sections S.3 - S.5.

4.2.2 Leading coefficient in MFPT series expansion

In Section 3.1, we have shown that for each $0 < \varepsilon < \varepsilon_0$, the unique stationary distribution $\pi(\varepsilon)$ for X^ε admits a real-analytic expansion in powers of ε. By (3.5) and (3.6), for $x \in \mathcal{X}$,

$$\frac{1}{\pi_x(\varepsilon)} = q_x(\varepsilon)\mathbb{E}_x[\zeta_x^\varepsilon] = 1 + \sum_{y \neq x} Q_{x,y}(\varepsilon) h_{y,x}(\varepsilon). \tag{4.15}$$

Recall that $E_0 = \{(x, y) : Q_{x,y}(\varepsilon) > 0 \text{ for all } \varepsilon \in (0, \varepsilon_{\max})\}$ and k_{xy} is the order of $Q_{x,y}(\varepsilon)$ for each $(x, y) \in E_0$. Using the algorithm in Section 4.2.1, we can obtain the order of the pole, $p_x(y)$, of the mean first passage time $h_{y,x}(\varepsilon)$ from y to x for all $y \neq x \in \mathcal{X}$. Therefore, for each $x \in \mathcal{X}$, the order of $\pi_x(\varepsilon)$ is

$$k_x = \max\{p_x(y) - k_{xy} : (x, y) \in E_0; 0\} \geq 0, \tag{4.16}$$

and then

$$\pi_x(\varepsilon) = \sum_{k=k_x}^{\infty} \varepsilon^k \pi_x^{(k)}.$$

The following theorem is for continuous time and builds on discrete time results of Avrachenkov et al. [5,6].

Theorem 4.7. *Suppose Assumptions 4.1, 4.2 and 4.5 hold. Let $Q_{\mathcal{A}}$ be given by (4.6), $\hat{X}_{\mathcal{A}}$ be as defined in (4.8), and α be the unique stationary distribution for $\hat{X}_{\mathcal{A}}$ defined in Theorem 4.2. Let $D = (-Q_{\mathcal{A}} + \mathbb{1}\alpha)^{-1} - \mathbb{1}\alpha$. For $y \in \mathcal{X}$, let k_y be the order of the stationary distribution $\pi_y(\varepsilon)$ of X^ε, defined by (4.16). Then, for $x, y \in \mathcal{A}$, the mean first passage time from x to y for X^ε is*

$$h_{x,y}(\varepsilon) = \frac{D_{y,y} - D_{x,y}}{\pi_y^{(k_y)}} \frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right). \tag{4.17}$$

Moreover, if $\hat{X}_{\mathcal{A}}$ is an irreducible Markov chain, then the order of the pole of $h_{x,y}(\varepsilon)$ is one, i.e., $k_y = 0$, and the coefficient of ε^{-1} in (4.17) is equal to the mean first passage time from x to y for the process $\hat{X}_{\mathcal{A}}$.

The proof of Theorem 4.7 is given in SI - Section S.7.1.

Remark 4.8. It may be possible that $D_{y,y} - D_{x,y} = 0$. In this case,

$$h_{x,y}(\varepsilon) = 0 \cdot \frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right) = O\left(\frac{1}{\varepsilon^{k_y}}\right).$$

However, if we find that the order of the pole of $h_{x,y}(\varepsilon)$ is k_y+1, using the algorithm in Section 4.2.1, then we can rule out the possibility of $D_{y,y} - D_{x,y}$ being *zero*.

4.2.3 Illustrative examples: 1D and 2D models We first apply the algorithm given in Section 4.2.1 to find the order of the pole of the time to memory loss in the 1D and 2D models introduced in Section 2. For the 1D model, we could also directly derive the analytical expression for the time to memory loss by exploiting first step analysis [26] and solve the system (3.2) introduced in Section 3.1 (see SI - Section S.8). Figure 5 illustrates the key steps of the algorithm for the 1D model, which lead to the conclusion that the time to memory loss for the active state is $\Theta(\varepsilon^{-1})$. Because of the symmetry in the input graph in Figure 5, the time to memory loss for the repressed state is also $\Theta(\varepsilon^{-1})$. These orders found by the algorithm are consistent with what can be directly derived by first step analysis. Similarly, SI - Figure S.1 illustrates the key steps of the algorithm for the 2D model, which leads to the conclusion that the time to memory loss of both the active and the repressed states is $\Theta(\varepsilon^{-1})$.

Next, we find the leading coefficient for the time to memory loss in the 1D and 2D models, which is the coefficient of the ε^{-1} term in all cases. Recall from Section 4.1.2 that Assumptions 4.1, 4.3 and 4.5 hold for both 1D and 2D models and hence by Lemma 4.4, so does Assumption 4.2 and $\hat{X}_{\mathcal{A}}$ is irreducible. For the 1D model, $Q_{\mathcal{A}}$ is given by (4.9). Thus, by Theorem 4.7, the leading coefficient of the time to memory loss for the active state is the mean first passage time

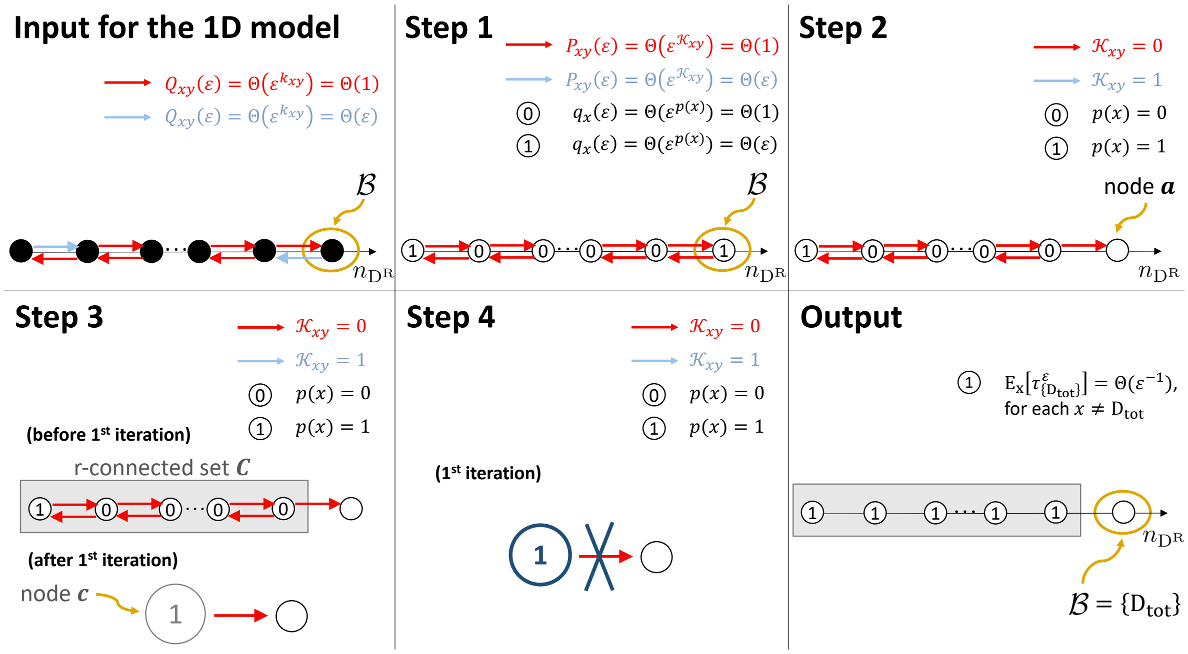

Figure 5: **Key steps of the algorithm for the 1D model.** The algorithm is described in Section 4.2.1, and it finds the order of the pole of the mean first passage time to $\mathcal{B} \subset \mathcal{X}$ from each state in $\mathcal{B}^c$. In our 1D model, the input for the algorithm is the order of each of the non-zero off-diagonal entries in $Q(\varepsilon)$ and the set $\mathcal{B} = \{\mathrm{D_{tot}}\}$. The order of the non-zero entries in $Q(\varepsilon)$ is represented by colored arrows in the graph in the "Input" panel. Step 1 transforms the orders in the $Q(\varepsilon)$-matrix into the orders in the $P(\varepsilon)$-matrix and the exponential parameters $q(\varepsilon)$ to give an equivalent construction for the continuous time Markov chain. The order of the non-zero entries in $P(\varepsilon)$ is represented by colored arrows in the graph, and the number in the circle at a state $x \in \mathcal{B}^c$ is the order of the pole $p(x)$ of $\frac{1}{q_x(\varepsilon)}$ (the mean sojourn time at the state x). In Step 2, the set $\mathcal{B}$ is relabeled as the node a, and then all transitions from a to $\mathcal{B}^c$ are removed. Step 3 for the 1D model involves only one iteration, where the collection of all nodes except the node a (called an r-connected set C) is condensed to a single node c, and the order of the pole at c is $p(c) = \max_{u \in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u,v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the 1^{st} iteration. Moreover, $\mathcal{K}_{ca} = \min\{\mathcal{K}_{ua} : u \in C \text{ and } (u,a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u,v) \in E\} = 0 - 0 = 0$. Step 4 involves one iteration. In this iteration, the node c is the only node other than a, so its value of p is fixed, and then any edges leading to or from c are removed. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

from the fully active state a to the fully repressed state r in $\hat{X}_{\mathcal{A}}$, which has an exponential distribution with parameter $(Q_{\mathcal{A}})_{a,r} = \frac{1-\mu}{1-\mu^{\mathrm{D_{tot}}}} \frac{k_E^A}{V} \mathrm{D}_{\mathrm{tot}}^2$ since $\hat{X}_{\mathcal{A}}$ has only two states. Thus,

$$h_{a,r}(\varepsilon) = \frac{1-\mu^{\mathrm{D_{tot}}}}{1-\mu} \frac{V}{k_E^A} \frac{1}{\mathrm{D}_{\mathrm{tot}}^2} \varepsilon^{-1} + O(1),$$

and similarly, the time to memory loss for the repressed state is

$$h_{r,a}(\varepsilon) = \frac{1-\mu^{\mathrm{D_{tot}}}}{1-\mu}\frac{V}{k_E^A}\frac{1}{b\mu^{\mathrm{D_{tot}}}\mathrm{D}_{\mathrm{tot}}^2}\varepsilon^{-1} + O(1).$$

Similarly, in the 2D model, by Theorem 4.7,

$$h_{a,r}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r}}\varepsilon^{-1} + O(1) \quad \text{and} \quad h_{r,a}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{r,a}}\varepsilon^{-1} + O(1).$$

As an example, when $\mathrm{D_{tot}} = 2$, $Q_{\mathcal{A}}$ is shown in (4.12) and we obtain that

$$h_{a,r}(\varepsilon) = \frac{V}{k_M^A}\frac{K}{4(k_{W0}^R + k_W^R)(k_{W0}^R + k_W^R + \frac{k_M^R}{V})}\varepsilon^{-1} + O(1) \quad \text{and}$$
$$h_{r,a}(\varepsilon) = \frac{V}{k_M^A}\frac{K}{4b\mu^2(k_{W0}^A + k_W^A)(k_{W0}^A + k_W^A + \frac{k_M^A}{V})}\varepsilon^{-1} + O(1),$$

with K defined in (4.13).

4.3 Monotonic dependence on parameters

An important aspect to consider in the study of the stochastic behavior of the chromatin modification circuit is that the erasure rate is different for each type of chromatin modification. These differences can introduce asymmetries in the system that can affect the stationary distribution and the time to memory loss of the active state and repressed state. These asymmetries are captured by the two parameters μ and μ'. In particular, μ quantifies the asymmetry between erasure rates of repressive and activating histone modifications and μ' quantifies the asymmetry between erasure rates of DNA methylation and activating histone modifications. In order to determine how the different chromatin modification erasure rates affect the stochastic behavior of the system, we study how μ and μ' affect the stationary distribution and the time to memory loss of the active and repressed gene states.

For the 1D model of the chromatin modification circuit, that does not include DNA methylation, we have an analytical expression for the stationary distribution and the time to memory loss ((2.5), (2.7), and (2.8)) and we can understand the effect of μ by directly studying the formulas. However, for the higher-than-1D models we do not have an explicit expression for the stationary distribution or time to memory loss. This is the reason why for these models we exploit the comparison theory developed in [13] that allows to determine how μ and μ' affect the stochastic behavior of the system through the construction of a coupling between processes with different values for these parameters. In the next subsection, we briefly summarize the relevant theory from [13].

4.3.1 Comparison theorems for continuous time Markov chains

Denote by $\leq$ the usual componentwise partial order on $\mathbb{R}^d$, i.e., for $x, y \in \mathbb{R}^d$, $x \leq y$ whenever $x_i \leq y_i$ for every $1 \leq i \leq d$. Let $m, d \geq 1$ be integers, consider a matrix $A \in \mathbb{R}^{m\times d}$, where no row of A is identically zero, and consider the following definition.

Definition 4.9 (Definition 3.1 from [13]). *For $x, y \in \mathbb{R}^d$, we say that $x \preccurlyeq_A y$ whenever $A(y - x) \geq 0$ and we say that $x \sim_A y$ whenever $Ax = Ay$.*

For the matrix A, consider the convex cone $K_A := \{x \in \mathbb{R}^d : \; Ax \geq 0\}$, and, for any $x \in \mathbb{R}^d$, consider the set $K_A + x = \{y \in \mathbb{R}^d : \; A(y-x) \geq 0\} = \{y \in \mathbb{R}^d : \; x \preccurlyeq_A y\}$ and the sets $\partial_i(K_A + x) := \{y \in K_A + x : \; \langle A_{i\bullet}, y\rangle = \langle A_{i\bullet}, x\rangle\}$ [3] for $1 \leq i \leq m$. Then, the boundary of $K_A + x$ can be expressed as

$$\partial(K_A + x) = \bigcup_{i=1}^{m} \partial_i(K_A + x).$$

Consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}^d_+$, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **increasing** with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $x \preccurlyeq_A y$ implies that $y \in \Gamma$. We observe that, for $x \in \mathcal{X}$, the set

$$(K_A + x) \cap \mathcal{X} = \{y \in \mathcal{X} : \; x \preccurlyeq_A y\} \tag{4.18}$$

is increasing by the transitivity property of $\preccurlyeq_A$. On the other hand, we will say that a set $\Gamma \subseteq \mathcal{X}$ is **decreasing** with respect to $\preccurlyeq_A$ if for every $x \in \Gamma$ and $y \in \mathcal{X}$, $y \preccurlyeq_A x$ implies that $y \in \Gamma$.

Now, consider a non-empty set $\mathcal{X} \subseteq \mathbb{Z}^d_+$ and a finite set of distinct nonzero possible transition vectors for a pair of continuous time Markov chains on $\mathcal{X}$. We denote the set of vectors by $\{v_1, \ldots, v_n\} \subseteq \mathbb{Z}^d \setminus \{0\}$, where 0 is the origin in $\mathbb{Z}^d$. Consider two collections of functions $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_n)$ and $\breve{\Upsilon} = (\breve{\Upsilon}_1, \ldots, \breve{\Upsilon}_n)$ from $\mathcal{X}$ into $\mathbb{R}_+$ such that $\Upsilon_j(x) = \breve{\Upsilon}_j(x) = 0$ if $x + v_j \notin \mathcal{X}$. Assume that $Q = (Q_{x,y})_{x,y\in\mathcal{X}}$, given by (3.11), is the infinitesimal generator for a continuous time Markov chain X and $\breve{Q}$, defined by (3.11) but with functions $\breve{\Upsilon}_1, \ldots, \breve{\Upsilon}_n$ in place of $\Upsilon_1, \ldots, \Upsilon_n$, is the infinitesimal generator for a continuous time Markov chain $\breve{X}$. We call X and $\breve{X}$ the continuous time Markov chains associated with Υ and $\breve{\Upsilon}$ respectively.

The following stochastic comparison result was proved in Campos et al. [13]. The condition (i) of the theorem and $A \in \mathbb{Z}^{n\times d}$ ensure that to go outside of $K_A + x$, the Markov chains will necessarily hit the boundary of $K_A + x$.

Theorem 4.10 (Theorems 3.2, 3.4, 3.5 from [13]). *With $\mathcal{X}$, $v_1, \ldots, v_n$, Υ and $\breve{\Upsilon}$ as described above, assume that the continuous time Markov chains associated with Υ and $\breve{\Upsilon}$ do not explode in finite time. Consider a matrix $A \in \mathbb{Z}^{m\times d}$ with nonzero rows and suppose that both of the following conditions hold:*

(i) For each $1 \leq j \leq n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.

(ii) For each $x \in \mathcal{X}$, $1 \leq i \leq m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that

$$\breve{\Upsilon}_j(y) \leq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j\rangle < 0, \tag{4.19}$$

and

$$\breve{\Upsilon}_j(y) \geq \Upsilon_j(x), \quad \text{for each } 1 \leq j \leq n \text{ such that } \langle A_{i\bullet}, v_j\rangle > 0. \tag{4.20}$$

[3] Here, for convenience of notation, let $A_{i\bullet}$ denote the row vector corresponding to the i-th row of A, for $1 \leq i \leq m$. In this article, we will adopt the convention of considering the inner product $\langle \cdot, \cdot \rangle$ as a function of a row vector in its first entry and as a function of a column vector in the second entry. In particular, $\langle A_{i\bullet}, x\rangle = \sum_{k=1}^{d} A_{ik} x_k$.

Then, for each pair $x^\circ, \breve{x}^\circ \in \mathcal{X}$ *such that* $x^\circ \preccurlyeq_A \breve{x}^\circ$, *there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with realizations of the two continuous time Markov chains* $X = \{X(t) : t \geq 0\}$ *and* $\breve{X} = \{\breve{X}(t) : t \geq 0\}$ *defined there, each having state space* $\mathcal{X} \subseteq \mathbb{Z}_+^d$, *with infinitesimal generators given by* Q *and* $\breve{Q}$, *associated with* Υ *and* $\breve{\Upsilon}$, *respectively, with initial conditions* $X(0) = x^\circ$ *and* $\breve{X}(0) = \breve{x}^\circ$, *and such that:*

$$\mathbb{P}\left[X(t) \preccurlyeq_A \breve{X}(t) \text{ for every } t \geq 0\right] = 1. \tag{4.21}$$

Furthermore, for a non-empty set $\Gamma \subseteq \mathcal{X}$, *consider* $\tau_\Gamma := \inf\{t \geq 0 : X(t) \in \Gamma\}$ *and* $\breve{\tau}_\Gamma := \inf\{t \geq 0 : \breve{X}(t) \in \Gamma\}$. *If* Γ *is increasing with respect to the relation* $\preccurlyeq_A$, *then* $\mathbb{E}[\breve{\tau}_\Gamma] \leq \mathbb{E}[\tau_\Gamma]$. *If* Γ *is decreasing with respect to the relation* $\preccurlyeq_A$, *then* $\mathbb{E}[\tau_\Gamma] \leq \mathbb{E}[\breve{\tau}_\Gamma]$. *Finally, suppose that the two continuous time Markov chains are irreducible and positive recurrent on* $\mathcal{X}$, *and denote the associated stationary distributions by* π *and* $\breve{\pi}$, *respectively. Then, if* $\Gamma \subseteq \mathcal{X}$ *is a non-empty set that is increasing with respect to* $\preccurlyeq_A$, *we have* $\sum_{x\in\Gamma} \pi_x \leq \sum_{x\in\Gamma} \breve{\pi}_x$, *or if* $\Gamma \subseteq \mathcal{X}$ *is a non-empty set that is decreasing with respect to* $\preccurlyeq_A$, *we have* $\sum_{x\in\Gamma} \breve{\pi}_x \leq \sum_{x\in\Gamma} \pi_x$.

4.3.2 Illustrative example: 2D model We are interested in determining how the asymmetry of the system, represented by the parameter $\mu = k_E^R/k_E^A$ affects the stationary distribution $\pi(\varepsilon)$ and the times to memory loss, $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, of the active ($a = (0, \mathrm{D_{tot}})^T$) and repressed ($r = (\mathrm{D_{tot}}, 0)^T$) states, respectively, for the continuous time Markov chain X^ε described in Section 2.2. For this, we use Theorem 4.10. For $\varepsilon \in (0, \varepsilon_0)$, let X^ε be the continuous time Markov chain with

$$\Upsilon_1(x) = f_A(x), \quad \Upsilon_2(x) = g_A^\varepsilon(x), \quad \Upsilon_3(x) = f_R(x), \quad \Upsilon_4(x) = g_R^\varepsilon(x), \quad x \in \mathcal{X} \tag{4.22}$$

with $\mathcal{X}$, $v_1, ..., v_4$, and $f_A(x)$, $g_A^\varepsilon(x)$, $f_R(x)$, $g_R^\varepsilon(x)$ as defined in Section 2.2, and introduce the continuous time Markov chain $\breve{X}^\varepsilon$ defined on $\mathcal{X}$, having the same transition vectors of X^ε, and having infinitesimal transition rates $\breve{\Upsilon}_1(x), ..., \breve{\Upsilon}_4(x)$ defined as for $\Upsilon_1(x), ..., \Upsilon_4(x)$, with all the parameters having the same values except that μ is replaced by $\breve{\mu}$, where $\mu \geq \breve{\mu}$. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{4.23}$$

and let us consider the partial order $x \preccurlyeq_A y$. A similar example was analyzed by Campos et al. [13] - Example 4.4, using the results of Theorem 4.10. The only differences are that, in [13], the matrix A is the negative of the matrix given in (4.23) and the inequality between μ and $\breve{\mu}$ is the opposite compared to the one considered here. The relationship between the notation in [13] and our notation is $\kappa_{1a} = k_{W0}^A + k_W^A$, $\kappa_{1b} = (k_M^A/V)$, $\kappa_{2a} = k_{W0}^R + k_W^R$, $\kappa_{2b} = (k_M^R/V)$, $\kappa_{3a} = \varepsilon(k_M^A/V)$, $\kappa_{3b} = (k_E^A/V)$, $c = b$.

From the analysis in [13], we can directly conclude that, if $\pi(\varepsilon)$ is the stationary distribution for X^ε and $\breve{\pi}(\varepsilon)$ the stationary distribution for $\breve{X}^\varepsilon$, then $\breve{\pi}_a(\varepsilon) \leq \pi_a(\varepsilon)$ and $\breve{\pi}_r(\varepsilon) \geq \pi_r(\varepsilon)$. This implies that increasing μ increases the probability of the system in steady-state being in the active state a to the detriment of the repressed state r (and vice versa for decreasing μ). We can also conclude, using natural notation for quantities associated with X^ε and $\breve{X}^\varepsilon$, that, defining $\tau_y^\varepsilon = \inf\{t \geq 0 : X^\varepsilon(t) = y\}$ and $\breve{\tau}_y^\varepsilon = \inf\{t \geq 0 : \breve{X}^\varepsilon(t) = y\}$, $h_{r,a}(\varepsilon) = \mathbb{E}_r[\tau_a^\varepsilon] \leq$

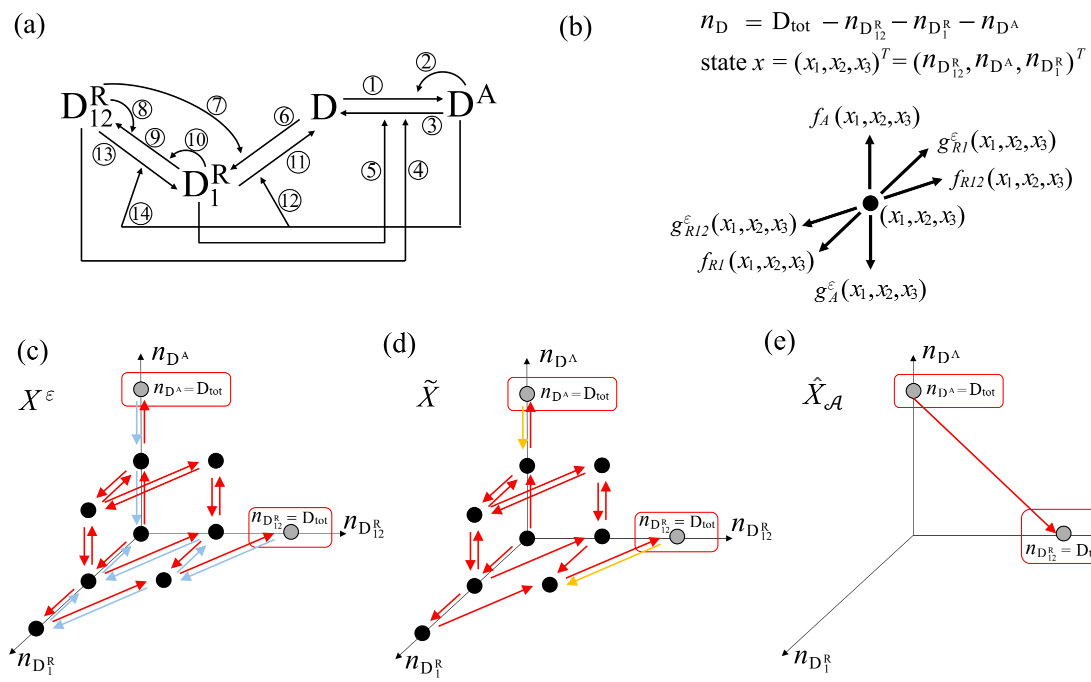

Figure 6: **3D Model and associated Markov chain.** (a) Original chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (5.1) in the main text. (b) Directions of the possible transitions of the Markov chain X^{ε}, starting from a state $x = (x_1, x_2, x_3)^T$ and whose rates are given in equation (5.2). (c) Graph for X^{ε}. Here, the red (blue) arrows correspond to $O(1)$ ($O(\varepsilon)$) transition rates. (d) Graph for the Markov chain $\tilde{X}$. Here, the gold arrows correspond to transitions that were $O(\varepsilon)$ in X^{ε} and became $O(1)$ in $\tilde{X}$. (e) Graph for the Markov chain $\hat{X}_{\mathcal{A}}$. For (c), (d), and (e) the state of the Markov chain is $x = (n_{\mathrm{D}^{\mathrm{R}}_{12}}, n_{\mathrm{D}^{\mathrm{A}}}, n_{\mathrm{D}^{\mathrm{R}}_{1}})^T$ and we consider $\mathrm{D_{tot}} = 2$. In panels (c) - (e), we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states.

$\mathbb{E}_r[\breve{\tau}^{\varepsilon}_a] = \breve{h}_{r,a}(\varepsilon)$ and $\breve{h}_{a,r}(\varepsilon) = \mathbb{E}_a[\breve{\tau}^{\varepsilon}_r] \leq \mathbb{E}_a[\tau^{\varepsilon}_r] = h_{a,r}(\varepsilon)$, implying that the time to memory loss of the repressed state decreases for higher values of μ, while the time to memory loss of the active state increases for higher values of μ.

5 Further Examples

In this section, compared to the models of the chromatin modification circuit introduced in Section 2, which do not include DNA methylation, we introduce more elaborate models that include DNA methylation and we study their stochastic behavior by exploiting the theory developed in this paper.

5.1 3D chromatin modification circuit model

We now introduce a model in which DNA methylation is also a possible chromatin mark. The species involved are D (unmodified nucleosome), $\mathrm{D}^{\mathrm{R}}_{1}$ (nucleosome with CpGme only), $\mathrm{D}^{\mathrm{R}}_{12}$ (nucleosome with both H3K9me3 and CpGme) and D^{A} (nucleosome with an activating histone modification). In particular, we assume that, in order to be modified with both repressive modifications, D is first modified with DNA methylation, obtaining $\mathrm{D}^{\mathrm{R}}_{1}$, and then with a repressive histone modification, obtaining

$\mathrm{D}^{\mathrm{R}}_{12}$. The opposite order of modifications is not allowed. This enables us to simplify the model and the related analysis. This assumption will be removed in the 4D model analyzed in Section 5.2. The chemical reaction system for the 3D model, shown in Fig. 6(a), is the following:

$$
\begin{aligned}
&① \ \mathrm{D} \xrightarrow{k^A_{W0} + k^A_W} \mathrm{D}^{\mathrm{A}}, \quad ② \ \mathrm{D} + \mathrm{D}^{\mathrm{A}} \xrightarrow{k^A_M} \mathrm{D}^{\mathrm{A}} + \mathrm{D}^{\mathrm{A}}, \quad ③ \ \mathrm{D}^{\mathrm{A}} \xrightarrow{\delta + \bar{k}^A_E} \mathrm{D},\\
&④ \ \mathrm{D}^{\mathrm{A}} + \mathrm{D}^{\mathrm{R}}_1 \xrightarrow{k^A_E} \mathrm{D} + \mathrm{D}^{\mathrm{R}}_1, ⑤ \ \mathrm{D}^{\mathrm{A}} + \mathrm{D}^{\mathrm{R}}_{12} \xrightarrow{2\,k^A_E} \mathrm{D} + \mathrm{D}^{\mathrm{R}}_{12},\\
(5.1) \quad &⑥ \ \mathrm{D} \xrightarrow{k^1_{W0} + k^1_W} \mathrm{D}^{\mathrm{R}}_1, \quad ⑦ \ \mathrm{D} + \mathrm{D}^{\mathrm{R}}_{12} \xrightarrow{k'_M} \mathrm{D}^{\mathrm{R}}_1 + \mathrm{D}^{\mathrm{R}}_{12}, \quad ⑧ \ \mathrm{D}^{\mathrm{R}}_1 + \mathrm{D}^{\mathrm{R}}_{12} \xrightarrow{k_M + \bar{k}_M} \mathrm{D}^{\mathrm{R}}_{12} + \mathrm{D}^{\mathrm{R}}_{12},\\
&⑨ \ \mathrm{D}^{\mathrm{R}}_1 \xrightarrow{k^2_{W0}} \mathrm{D}^{\mathrm{R}}_{12}, \quad ⑩ \ \mathrm{D}^{\mathrm{R}}_1 + \mathrm{D}^{\mathrm{R}}_1 \xrightarrow{\bar{k}_M} \mathrm{D}^{\mathrm{R}}_{12} + \mathrm{D}^{\mathrm{R}}_1, ⑪, \quad \mathrm{D}^{\mathrm{R}}_1 \xrightarrow{\delta' + k'_T} \mathrm{D},\\
&⑫ \ \mathrm{D}^{\mathrm{R}}_1 + \mathrm{D}^{\mathrm{A}} \xrightarrow{k'^*_T} \mathrm{D} + \mathrm{D}^{\mathrm{A}}, ⑬ \ \mathrm{D}^{\mathrm{R}}_{12} \xrightarrow{\delta + \bar{k}^R_E} \mathrm{D}^{\mathrm{R}}_1, \quad ⑭ \ \mathrm{D}^{\mathrm{R}}_{12} + \mathrm{D}^{\mathrm{A}} \xrightarrow{k^R_E} \mathrm{D}^{\mathrm{R}}_1 + \mathrm{D}^{\mathrm{A}},
\end{aligned}
$$

where $k^A_{W0}, k^A_W, k^A_M, \delta, \bar{k}^A_E, k^A_E, k^1_{W0}, k^1_W, k^2_{W0}, k^2_W, k'_M, \bar{k}_M, k_M, \delta', k'_T, k'^*_T, \bar{k}^R_E, k^R_E > 0$ and the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. As we did for the 2D model, define parameters $\varepsilon = \frac{\delta + \bar{k}^A_E}{\frac{k^A_M}{V}\mathrm{D_{tot}}}$ and $\mu = \frac{k^R_E}{k^A_E}$, with a constant b such that $\frac{\delta + \bar{k}^R_E}{\delta + \bar{k}^A_E} = b\mu$. Furthermore, since this model includes DNA methylation, we also define $\mu' = \frac{k'^*_T}{k^A_E}$ and a constant β such that $\frac{\delta' + k'_T}{\delta + \bar{k}^A_E} = \beta\mu'$. The parameter μ' quantifies the asymmetry between the erasure rates of DNA methylation and activating histone modifications. The Markov chain X^ε associated with the system is a linearly perturbed finite state continuous time Markov chain with the state x tracking $n_{\mathrm{D}^{\mathrm{R}}_{12}}$, $n_{\mathrm{D}^{\mathrm{A}}}$, $n_{\mathrm{D}^{\mathrm{R}}_1}$, that is, the number of nucleosomes of types $\mathrm{D}^{\mathrm{R}}_{12}$, D^{A}, and $\mathrm{D}^{\mathrm{R}}_1$, respectively. If the total number of modifiable nucleosomes is $\mathrm{D_{tot}}$, which is conserved, the state space is $\mathcal{X} = \{(x_1, x_2, x_3)^T \in \mathbb{Z}^3_+ : x_1 + x_2 + x_3 \leq \mathrm{D_{tot}}\}$. The transition vectors for X^ε are given by $v_1 = -v_2 = (0,1,0)^T$, $v_3 = -v_4 = (0,0,1)^T$, and $v_5 = -v_6 = (1,0,-1)^T$. The infinitesimal transition rates are

$$
(5.2) \quad
\begin{aligned}
Q_{x,x+v_1}(\varepsilon) &= f_A(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3))\left(k^A_{W0} + k^A_W + \frac{k^A_M}{V}x_2\right),\\
Q_{x,x+v_2}(\varepsilon) &= g^\varepsilon_A(x) = x_2\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}} + \frac{k^A_E}{V}(x_3 + 2x_1)\right),\\
Q_{x,x+v_3}(\varepsilon) &= f_{R1}(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3))\left(k^1_{W0} + k^1_W + \frac{k'_M}{V}x_1\right),\\
Q_{x,x+v_4}(\varepsilon) &= g^\varepsilon_{R1}(x) = x_3\mu'\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}\beta + x_2\frac{k^A_E}{V}\right),\\
Q_{x,x+v_5}(\varepsilon) &= f_{R12}(x) = x_3\left(k^2_{W0} + \frac{k_M}{V}x_1 + \frac{\bar{k}_M}{V}\left(x_1 + \frac{x_3 - 1}{2}\right)\right),\\
Q_{x,x+v_6}(\varepsilon) &= g^\varepsilon_{R12}(x) = x_1\mu\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}b + x_2\frac{k^A_E}{V}\right).
\end{aligned}
$$

A representation of the possible transitions, with associated rates, and the Markov chain graph for $\mathrm{D_{tot}} = 2$ are given in Fig. 6(b) and (c), respectively. Each rate depends on the state x.

5.1.1 Stationary distribution We now focus on the expansion as a function of ε of the stationary distribution for the 3D model. In SI - Section S.10, we show that, when $\varepsilon = 0$, the continuous time Markov chain associated with the 3D model has transient states $\mathcal{T} = \{i_1, \dots, i_m\}$ where $m = \sum_{j=0}^{\mathrm{D_{tot}}} \left(\frac{(j+2)(j+1)}{2}\right) - 2$, $i_1 = (0, \mathrm{D_{tot}} - 1, 0)^T$, $i_m = (\mathrm{D_{tot}} - 1, 0, 1)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, \mathrm{D_{tot}}, 0)^T$ corresponding to the fully active state ($n_{\mathrm{D^A}} = \mathrm{D_{tot}}$) and $r = (\mathrm{D_{tot}}, 0, 0)^T$ corresponding to the fully repressed state ($n_{\mathrm{D^R_{12}}} = \mathrm{D_{tot}}$), respectively. Then, Assumption 4.1 holds (see SI - Section S.10). Furthermore, $\mathcal{X} = \mathcal{A} \cup \mathcal{T}$ and from (5.2) we see that $Q(\varepsilon)$ can be written in the form (4.2), where $Q(\varepsilon)$ is a linear perturbation of $Q(0)$. Hence, Assumption 4.5 holds. Assumption 4.2 also holds, where the recurrent class is $\{r\}$ (see SI - Section S.10). Then, we can apply SI - Theorem S.9. We first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0 \dots, 0]$ where α is the unique stationary distribution for the process $\hat{X}_{\mathcal{A}}$ with infinitesimal generator $Q_{\mathcal{A}} = A_1 + S_1(-T_0)^{-1}R_0$. Since the recurrent class $\{r\}$ is a singleton and α is supported on $\{r\}$, we must have $\alpha_a = 0$ and $\alpha_r = 1$.

We now derive an expression for $\pi^{(1)}$. For the transient states $\mathcal{T} = \{i_1, \dots, i_m\}$, $\beta^{(1)} = [\pi^{(1)}_{i_1}, ..., \pi^{(1)}_{i_m}] = \alpha S_1(-T_0)^{-1} = [0, ..., 0, \pi^{(1)}_{i_m}]$, with

$$\pi^{(1)}_{i_m} = \frac{\mu b \frac{k_M^A}{V} \mathrm{D_{tot}^2}}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D_{tot}} - 1)}.$$

See SI - Section S.10 for the detailed mathematical derivation. Now, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$, $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$.

As an illustration, suppose $\mathrm{D_{tot}} = 2$. Then (see SI - Section S.10 for the detailed mathematical derivation),

$$Q_{\mathcal{A}} = \frac{K_1 + \mu K_2}{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \tag{5.3}$$

$$\pi_a^{(1)} = \frac{\mu^2 \mu'^2 K_7}{K_8(K_9 + K_{10}\mu)}, \quad \pi_r^{(1)} = -\pi_a^{(1)} - \pi_{i_m}^{(1)} = -\frac{\mu^2\mu'^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)}, \tag{5.4}$$

with $m = 8$ and K_i, $i = 1, ..., 11$, are non-negative constants independent of ε, μ and μ' (see SI - Section S.10 for their precise definitions). Hence, the stationary distribution for $\mathrm{D_{tot}} = 2$ satisfies

$$\pi_x(\varepsilon) = \begin{cases} \varepsilon \frac{(\mu\mu')^2 K_7}{K_8(K_9 + K_{10}\mu)} + O(\varepsilon^2) & \text{if } x = a = (0,2,0)^T \\ O(\varepsilon^2) & \text{if } x \in \mathcal{T} \backslash \{i_m\} \\ \varepsilon \frac{K_{11}}{K_8}\mu + O(\varepsilon^2) & \text{if } x = i_m = (1,0,1)^T \\ 1 - \varepsilon \frac{(\mu\mu')^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)} + O(\varepsilon^2) & \text{if } x = r = (2,0,0)^T. \end{cases} \tag{5.5}$$

For small $\varepsilon > 0$, the stationary distribution is concentrated around the active and repressed states, although more mass is concentrated around the repressed state. However, higher values of μ' increase the probability of being in the active state, while decreasing the probability of being in the repressed state.

5.1.2 Time to memory loss In this section, we determine how the leakage of the system (ε) and the asymmetry between activating histone modifications and DNA methylation (μ') affect the time to memory loss of the active state $h_{a,r}(\varepsilon)$ and the time to memory loss of the repressed state $h_{r,a}(\varepsilon)$.

Firstly, by the algorithm in Section 4.2.1, we have that $h_{a,r}(\varepsilon)$ is $O(\varepsilon^{-1})$ and $h_{r,a}(\varepsilon)$ is $O(\varepsilon^{-2})$ (see SI - Section S.6). This means that decreasing the leakage extends the memory of both the active and repressed chromatin states, but the effect is stronger for the repressed state. This difference is influenced by the co-existence and cooperation between DNA methylation and repressive histone marks that introduce a structural bias in the 3D chromatin modification circuit towards a repressed chromatin state.

These results are consistent with the ones obtained by applying Theorem 4.7, which allows us not only to find the order of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, but also to find an expression for their leading coefficients (see SI - Section S.10 for the detailed mathematical derivation). As an example, when $\mathrm{D_{tot}} = 2$, $Q_{\mathcal{A}}$ and $\pi_a^{(1)}$ are shown in (5.3) and (5.4), and we obtain from them that

$$h_{a,r}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6}{K_1 + \mu K_2}\frac{1}{\varepsilon} + O(1), \quad \text{and} \tag{5.6}$$

$$h_{r,a}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6}{K_1 + \mu K_2}\frac{K_8(K_9 + K_{10}\mu)}{\mu^2\mu'^2 K_7}\frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right), \tag{5.7}$$

where K_i, $i = 1, ..., 11$, are non-negative constants independent of ε, μ and μ', defined in SI - Section S.10.

Now, we focus on understanding how the asymmetry between chromatin modification erasure rates affects the time to memory loss. In particular, since experimental data suggest that the asymmetry between the erasure rates of DNA methylation and activating histone modifications is more pronounced than the asymmetry between erasure rates of opposite histone modifications, in this analysis we focus only on studying the effect of μ', but a similar procedure to the one presented in the next paragraph could be applied to study the effect of μ. To this end, we exploit the comparison Theorem 4.10 to determine directly how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, without deriving an explicit expression for them. To this end, we first note that the transitions of the Markov chain $X^\varepsilon(t)$ are in six possible directions, that can be written as $v_1 = (0,1,0)^T$, $v_2 = (0,-1,0)^T$, $v_3 = (0,0,1)^T$, $v_4 = (0,0,-1)^T$, $v_5 = (1,0,-1)^T$, $v_6 = (-1,0,1)^T$, with the associated infinitesimal transition rates that can be written as $\Upsilon_1(x) = f_A(x)$, $\Upsilon_2(x) = g_\varepsilon^A(x)$, $\Upsilon_3(x) = f_{R1}(x)$, $\Upsilon_4(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_5(x) = f_{R12}(x)$, $\Upsilon_6(x) = g_{R12}^\varepsilon(x)$. Define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and, for $x \in \mathcal{X}$, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preccurlyeq_A w\}$. Let us also introduce infinitesimal transition rates $\breve{\Upsilon}_i(x)$, $i = 1, 2, ..., 6$, defined as for $\Upsilon_i(x)$, $i = 1, 2, ..., 6$, with all the parameters having the same values except that μ' is replaced by $\breve{\mu}'$, with $\mu' \geq \breve{\mu}'$. All of the conditions of Theorem 4.10 hold (see SI - Section S.10) and so we can apply the theorem. This allows us to establish that, since $a = (0, \mathrm{D_{tot}}, 0)^T \preccurlyeq_A r = (\mathrm{D_{tot}}, 0, 0)^T$, then $\breve{h}_{a,r}(\varepsilon) \leq h_{a,r}(\varepsilon)$ and

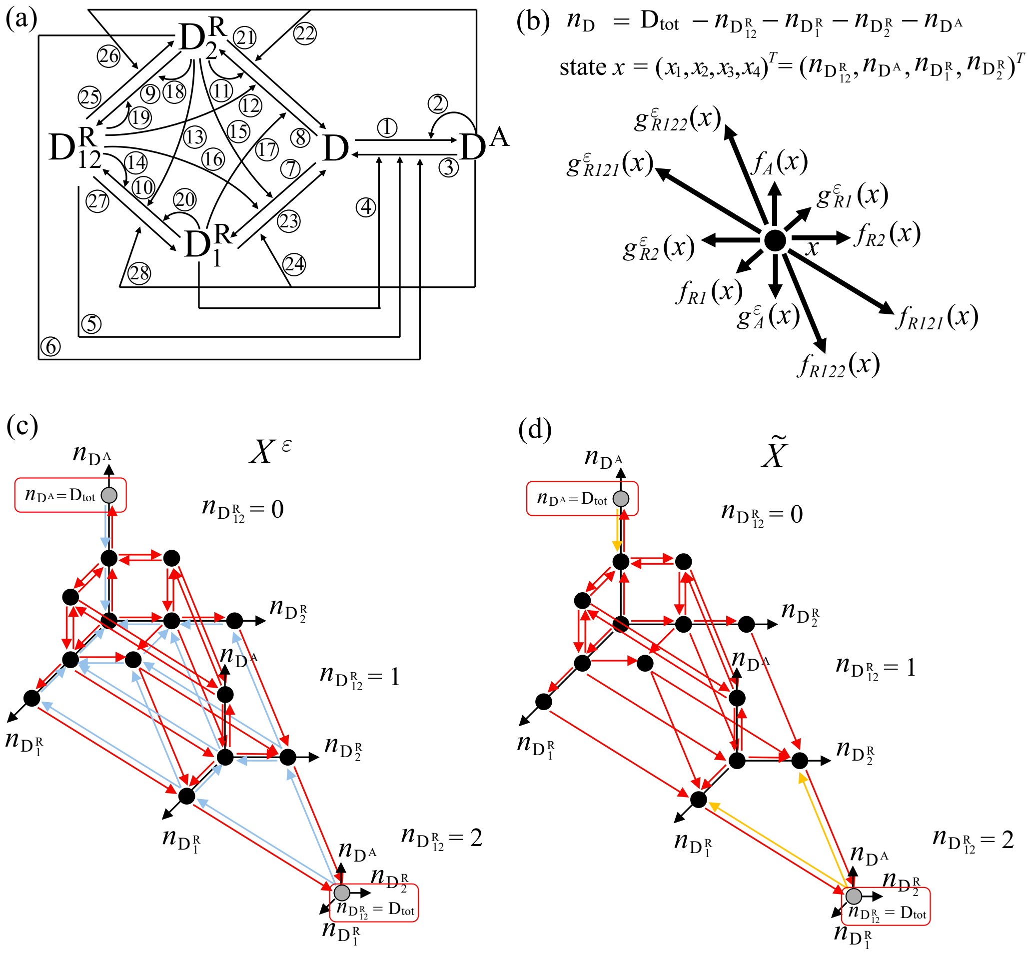

Figure 7: **4D Model and associated Markov chain.** (a) Chemical reaction system. The numbers on the arrows correspond to the reactions associated with the arrows as described in (5.8) in the main text. (b) Directions of the possible transitions of the Markov chain X^ε associated with the reduced SCRN, starting from a state $x = (x_1, x_2, x_3, x_4)^T$ and whose rates are given in equation (5.9). (c) Graph for X^ε. Here, the red (blue) arrows correspond to $O(1)$ ($O(\varepsilon)$) transition rates. (d) Graph for the Markov chain $\tilde{X}$. Here, the golden arrows correspond to the transitions that were $O(\varepsilon)$ in X^ε and became $O(1)$ in $\tilde{X}$. For (c) and (d) the state of the Markov chain is $x = (n_{D^R_{12}}, n_{D^A}, n_{D^R_1}, n_{D^R_2})^T$, we consider $D_{tot} = 2$, and we show three interconnected slices ($n_{D^R_{12}} = 0, 1, 2$) of the Markov chain state space. In panels (c) and (d), we use gray dots to represent the states belonging to $\mathcal{A}$ and black dots to represent all the other states.

$h_{r,a}(\varepsilon) \leq \breve{h}_{r,a}(\varepsilon)$, where $\breve{}$ indicates quantities associated with $\breve{\Upsilon}$. Thus, we can conclude that, given that the only difference between the two systems was that $\mu' \geq \breve{\mu}'$, the time to memory loss of the active state is monotonically increasing with μ', while the time to memory loss of the repressed state is monotonically decreasing with μ'.

5.2 4D chromatin modification circuit model Now, we consider a complete model in which the species involved are D, D_1^R, D_{12}^R, D^A and D_2^R (nucleosome with H3K9me3 only). Compared to the 3D model, we assume that, in order to be modified with both repressive modifications, D can be also modified first with a repressive histone modification (H3K9me3), obtaining D_2^R, and then with DNA methylation (CpGme), obtaining D_{12}^R. The chemical reaction system, shown in Fig. 7(a), is the following:

$$
\begin{aligned}
&\text{①}\ D \xrightarrow{k_{W0}^A + k_W^A} D^A,\quad \text{②}\ D + D^A \xrightarrow{k_M^A} D^A + D^A,\quad \text{③}\ D^A \xrightarrow{\delta + \bar{k}_E^A} D,\\
&\text{④}\ D^A + D_1^R \xrightarrow{k_E^A} D + D_1^R,\quad \text{⑤}\ D^A + D_{12}^R \xrightarrow{2\,k_E^A} D + D_{12}^R,\quad \text{⑥}\ D^A + D_2^R \xrightarrow{k_E^A} D + D_2^R,\\
&\text{⑦}\ D \xrightarrow{k_{W0}^1 + k_W^1} D_1^R,\quad \text{⑧}\ D \xrightarrow{k_{W0}^2 + k_W^2} D_2^R,\quad \text{⑨}\ D_2^R \xrightarrow{k_{W0}^1} D_{12}^R,\quad \text{⑩}\ D_1^R \xrightarrow{k_{W0}^2} D_{12}^R,\\
&\text{⑪}\ D + D_2^R \xrightarrow{k_M} D_2^R + D_2^R,\quad \text{⑫}\ D + D_{12}^R \xrightarrow{k_M + \bar{k}_M} D_2^R + D_{12}^R,\\
(5.8)\quad &\text{⑬}\ D_1^R + D_2^R \xrightarrow{k_M} D_{12}^R + D_2^R,\quad \text{⑭}\ D_1^R + D_{12}^R \xrightarrow{k_M + \bar{k}_M} D_{12}^R + D_{12}^R,\\
&\text{⑮}\ D + D_2^R \xrightarrow{k_M^{'}} D_1^R + D_2^R,\quad \text{⑯}\ D + D_{12}^R \xrightarrow{k_M^{'}} D_1^R + D_{12}^R,\quad \text{⑰}\ D + D_1^R \xrightarrow{\bar{k}_M} D_2^R + D_1^R\\
&\text{⑱}\ D_2^R + D_2^R \xrightarrow{k_M^{'}} D_{12}^R + D_2^R,\quad \text{⑲}\ D_2^R + D_{12}^R \xrightarrow{k_M^{'}} D_{12}^R + D_{12}^R,\\
&\text{⑳}\ D_1^R + D_1^R \xrightarrow{\bar{k}_M} D_{12}^R + D_1^R,\quad \text{㉑}\ D_2^R \xrightarrow{\delta + \bar{k}_E^R} D,\quad \text{㉒}\ D_2^R + D^A \xrightarrow{k_E^R} D + D^A,\\
&\text{㉓}\ D_1^R \xrightarrow{\delta^{'} + k_T^{'}} D\quad \text{㉗}\ D_1^R + D^A \xrightarrow{k_T^{'*}} D + D^A,\quad \text{㉔}\ D_{12}^R \xrightarrow{\delta^{'} + k_T^{'}} D_2^R,\\
&\text{㉖}\ D_{12}^R + D^A \xrightarrow{k_T^{'*}} D_2^R + D^A, \text{㉗}\ D_{12}^R \xrightarrow{\delta + \bar{k}_E^R} D_1^R,\quad \text{㉘}\ D_{12}^R + D^A \xrightarrow{k_E^R} D_1^R + D^A,
\end{aligned}
$$

in which the form of the reaction rate constants is due to the fact that reactions with the same reactants and products have been combined. As we did for the 3D model, let us define the parameter $\varepsilon = \frac{\delta+\bar{k}_E^A}{k_M^A \frac{D_{tot}}{V}}$, the parameter $\mu = \frac{k_E^R}{k_E^A}$, with a constant b such that $\frac{\delta+\bar{k}_E^R}{\delta+\bar{k}_E^A} = b\mu$, and the parameter $\mu' = \frac{k_T^{'*}}{k_E^A}$, with a constant β such that $\frac{\delta^{'}+k_T^{'}}{\delta+\bar{k}_E^A} = \beta\mu'$. The Markov chain X^ε associated with the system is a linearly perturbed finite state continuous time Markov chain with the state x representing the number of each type of modified nucleosome, i.e., $x = (n_{D_{12}^R}, n_{D^A}, n_{D_1^R}, n_{D_2^R})^T = (x_1, x_2, x_3, x_4)^T$. If the total number of nucleosomes is D_{tot}, which is conserved, then the state space is $\mathcal{X} = \{(x_1, x_2, x_3, x_4)^T \in \mathbb{Z}_+^4 : x_1+x_2+x_3+x_4 \leq D_{tot}\}$. The transition vectors for X^ε are given by $v_1 = (0,1,0,0)^T$, $v_2 = (0,-1,0,0)^T$, $v_3 = (0,0,1,0)^T$, $v_4 = (0,0,-1,0)^T$, $v_5 = (0,0,0,1)^T$, $v_6 = (0,0,0,-1)^T$, $v_7 = (1,0,-1,0)^T$, $v_8 = (-1,0,1,0)^T$, $v_9 = (1,0,0,-1)^T$ and $v_{10} = (-1,0,0,1)^T$. The infinitesimal transition rates are

$$
Q_{x,x+v_1}(\varepsilon) = f_A(x) = (D_{tot} - (x_1+x_2+x_3+x_4))\left(k_{W0}^A + k_W^A + \frac{k_M^A}{V}x_2\right),
$$

$$
Q_{x,x+v_2}(\varepsilon) = g_A^\varepsilon(x) = x_2\left(\varepsilon\frac{k_M^A}{V}D_{tot} + \frac{k_E^A}{V}(x_3+x_4+2x_1)\right),
$$

$$
Q_{x,x+v_3}(\varepsilon) = f_{R1}(x) = (D_{tot} - (x_1+x_2+x_3+x_4))\left(k_{W0}^1 + k_W^1 + \frac{k_M^{'}}{V}(x_1+x_4)\right),
$$

$$Q_{x,x+v_4}(\varepsilon) = g^{\varepsilon}_{R1}(x) = x_3\mu'\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}\beta + x_2\frac{k^A_E}{V}\right),$$

$$Q_{x,x+v_5}(\varepsilon) = f_{R2}(x) = (\mathrm{D_{tot}} - (x_1+x_2+x_3+x_4))\left(k^2_{W0} + k^2_W + \frac{k_M}{V}(x_1+x_4) + \frac{\bar{k}_M}{V}(x_1+x_3)\right),$$

(5.9)

$$Q_{x,x+v_6}(\varepsilon) = g^{\varepsilon}_{R2}(x) = x_4\mu\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}b + x_2\frac{k^A_E}{V}\right),$$

$$Q_{x,x+v_7}(\varepsilon) = f_{R121}(x) = x_3\left(k^2_{W0} + \frac{k_M}{V}(x_1+x_4) + \frac{\bar{k}_M}{V}\left(x_1 + \frac{x_3-1}{2}\right)\right),$$

$$Q_{x,x+v_8}(\varepsilon) = g^{\varepsilon}_{R121}(x) = x_1\mu\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}b + x_2\frac{k^A_E}{V}\right),$$

$$Q_{x,x+v_9}(\varepsilon) = f_{R122}(x) = x_4\left(k^1_{W0} + \frac{k'_M}{V}\left(x_1 + \frac{x_4-1}{2}\right)\right),$$

$$Q_{x,x+v_{10}}(\varepsilon) = g^{\varepsilon}_{R122}(x) = x_1\mu'\left(\varepsilon\frac{k^A_M}{V}\mathrm{D_{tot}}\beta + x_2\frac{k^A_E}{V}\right).$$

A representation of the transition vectors and the Markov chain graph for $\mathrm{D_{tot}} = 2$ are given in Fig. 7 (b) and (c), respectively. As before, each rate depends on the state x, but in the rest of the section we will not show this dependency to simplify the notation. Now, we determine the stochastic behavior of the full chromatin modification circuit model in terms of its stationary distribution and time to memory loss. For this study, we will consider $k^A_W = k^1_W = k^2_W = 0$ (i.e., there are no external transcription factors enhancing the establishment of chromatin modifications). This assumption will not change the qualitative nature of the results focused on studying the effect of ε, μ, and μ' on the stochastic behavior of the chromatin modification circuit model.

5.2.1 Stationary distribution We now determine the zeroth and first order terms of the stationary distribution expansion for the 4D model. As shown in SI - Section S.11, when $\varepsilon = 0$, the continuous time Markov chain associated with the 4D model has transient states $\mathcal{T} = \{i_1, \ldots, i_m\}$ where $m = \sum_{j=0}^{\mathrm{D_{tot}}}\sum_{k=0}^{j}\left(\frac{(k+2)(k+1)}{2}\right) - 2$, $i_1 = (0, \mathrm{D_{tot}}-1, 0, 0)^T$, $i_{m-1} = (\mathrm{D_{tot}} - 1, 0, 0, 1)^T$, $i_m = (\mathrm{D_{tot}}-1, 0, 1, 0)^T$, and absorbing states $\mathcal{A} = \{a, r\}$, with $a = (0, \mathrm{D_{tot}}, 0, 0)^T$ corresponding to the fully active state ($n_{\mathrm{D^A}} = \mathrm{D_{tot}}$) and $r = (\mathrm{D_{tot}}, 0, 0, 0)^T$ corresponding to the fully repressed state ($n_{\mathrm{D^R_{12}}} = \mathrm{D_{tot}}$), respectively. Then, Assumption 4.1 holds (see SI - Section S.11), so that $\mathcal{X} = \mathcal{A} \cup \mathcal{T}$, and we can rewrite the infinitesimal generator $Q(\varepsilon)$ in the form of (4.2) (see SI - Section S.11), where the perturbation is linear and so Assumption 4.5 holds. We can also verify Assumption 4.2 (see SI - Section S.11). Hence, we can apply SI - Theorem S.9, as was done for the 3D model.

In particular, we obtain $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0 \ldots, 0]$, with $\alpha_a = 0$, $\alpha_r = 1$, and $\beta^{(1)} = [\pi^{(1)}_{i_1}, ..., \pi^{(1)}_{i_m}] = [0, ..., 0, \pi^{(1)}_{i_{m-1}}, \pi^{(1)}_{i_m}]$, with

$$\pi^{(1)}_{i_{m-1}} = \frac{\mu'\beta\frac{k^A_M}{V}\mathrm{D^2_{tot}}}{k^1_{W0} + \frac{k'_M}{V}(\mathrm{D_{tot}}-1)}, \qquad \pi^{(1)}_{i_m} = \frac{\mu b\frac{k^A_M}{V}\mathrm{D^2_{tot}}}{k^2_{W0} + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D_{tot}}-1)}.$$

See SI - Section S.11 for the detailed mathematical derivation. Now, $\alpha^{(1)} = [\pi^{(1)}_a, \pi^{(1)}_r]$ is the unique vector such that $\alpha^{(1)}Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$, $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$.

As an example, suppose $\mathrm{D_{tot}} = 2$ and assume $\beta = b$, $k^1_{W0} = k^2_{W0} = k^A_{W0}$ and $k'_M = \bar{k}_M = k_M = k^A_M$. These assumptions do not affect the final qualitative conclusions related to the effect of ε, μ and μ' on the stationary distribution. Then (see SI - Section S.11 for the detailed mathematical derivation)

$$Q_{\mathcal{A}} = \frac{\bar{K}_1(\mu,\mu')}{\bar{K}_2(\mu,\mu')} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \tag{5.10}$$

$$\pi^{(1)}_a = \frac{\bar{K}_3(\mu,\mu')}{\bar{K}_4(\mu,\mu')}, \quad \pi^{(1)}_r = -\pi^{(1)}_a - \pi^{(1)}_{i_{m-1}} - \pi^{(1)}_{i_m} = -\frac{\bar{K}_3(\mu,\mu')}{\bar{K}_4(\mu,\mu')} - \mu' K_{18} - \mu K_{19}, \tag{5.11}$$

with

$$\begin{aligned} \bar{K}_1(\mu,\mu') &= K_1((\mu')^2 K_2 + (\mu)^2 K_3 + \mu\mu' K_4 + \mu' K_5 + \mu K_6 + K_7), \\ \bar{K}_2(\mu,\mu') &= \mu'\mu(\mu'+\mu)K_8 + (\mu')^2 K_9 + (\mu)^2 K_{10} + \mu\mu' K_{11} + \mu' K_{12} + \mu K_{13} + K_{14}, \\ \bar{K}_3(\mu,\mu') &= (\mu\mu')^2 K_{15}((\mu+\mu')K_{16} + K_{17}), \\ \bar{K}_4(\mu,\mu') &= K_{20}((\mu')^2 K_2 + (\mu)^2 K_3 + \mu\mu' K_4 + \mu' K_5 + \mu K_6 + K_7), \end{aligned} \tag{5.12}$$

in which $m = 13$ and K_i, $i = 1, ..., 20$, are non-negative functions independent of μ and μ' (see SI - Section S.11 for their precise definitions). We then have

$$\pi_x(\varepsilon) = \begin{cases} \varepsilon\frac{\bar{K}_3(\mu,\mu')}{\bar{K}_4(\mu,\mu')} + O(\varepsilon^2) & \text{if } x = a = (0,2,0,0)^T \\ O(\varepsilon^2) & \text{if } x \in \mathcal{T}\backslash\{i_{m-1}, i_m\} \\ \varepsilon\mu' K_{18} + O(\varepsilon^2) & \text{if } x = i_{m-1} = (1,0,0,1)^T \\ \varepsilon\mu K_{19} + O(\varepsilon^2) & \text{if } x = i_m = (1,0,1,0)^T \\ 1 - \varepsilon\left(\frac{\bar{K}_3(\mu,\mu')}{\bar{K}_4(\mu,\mu')} + \mu' K_{18} + \mu K_{19}\right) + O(\varepsilon^2) & \text{if } x = r = (2,0,0,0)^T. \end{cases}$$

For small $\varepsilon > 0$, the stationary distribution is concentrated around the active and repressed states, and higher values of μ' or μ shift the distribution towards the active state.

5.2.2 Time to memory loss

As was done for the 3D model, we determine for the 4D model how the parameters ε and μ' affect the time to memory loss of the active state, $h_{a,r}(\varepsilon)$, and the time to memory loss of the repressed state, $h_{r,a}(\varepsilon)$. Firstly, by the algorithm in Section 4.2.1, $h_{a,r}(\varepsilon)$ is $O(\varepsilon^{-1})$ and $h_{r,a}(\varepsilon)$ is $O(\varepsilon^{-2})$ (see SI - Section S.6). Then, by applying Theorem 4.7 we can obtain expressions for the leading coefficients of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$ (See SI - Section S.11 for the detailed mathematical derivation). As an example, when $\mathrm{D_{tot}} = 2$, $Q_{\mathcal{A}}$ and $\pi^{(1)}_a$ are shown in (5.10) and (5.11), and we obtain that

$$h_{a,r}(\varepsilon) = \frac{\bar{K}_2(\mu,\mu')}{\bar{K}_1(\mu,\mu')}\frac{1}{\varepsilon} + O(1), \quad \text{and} \quad h_{r,a}(\varepsilon) = \frac{\bar{K}_2(\mu,\mu')K_4(\mu,\mu')}{\bar{K}_1(\mu,\mu')\bar{K}_3(\mu,\mu')}\frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right),$$

where $\bar{K}_i(\mu,\mu')$, $i = 1, ..., 4$, are defined in (5.12).

Now, we determine how μ', the parameter encapsulating the asymmetry between the DNA methylation erasure rate and the activating histone modification erasure rate, affects the time to memory loss. To this end, we seek to determine directly how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, without deriving an explicit expression for them. To this end, we would like to exploit two theorems from [13], namely, Theorem S.2 and Theorem 3.4 there. The transitions of the Markov chain X^ε are in ten possible directions, $v_1 = -v_2 = (1,0,-1,0)^T$, $v_3 = -v_4 = (1,0,0,-1)^T$, $v_5 = -v_6 = (0,1,0,0)^T$, $v_7 = -v_8 = (0,0,1,0)^T$, and $v_9 = -v_{10} = (0,0,0,1)^T$, with the associated infinitesimal transition rates $\Upsilon_1(x) = f_{R121}(x)$, $\Upsilon_2(x) = g^\varepsilon_{R121}(x)$, $\Upsilon_3(x) = f_{R122}(x)$, $\Upsilon_4(x) = g^\varepsilon_{R122}(x)$, $\Upsilon_5(x) = f_A(x)$, $\Upsilon_6(x) = g^\varepsilon_A(x)$, $\Upsilon_7(x) = f_{R1}(x)$, $\Upsilon_8(x) = g^\varepsilon_{R1}(x)$, $\Upsilon_9(x) = f_{R2}(x)$, $\Upsilon_{10}(x) = g^\varepsilon_{R2}(x)$. Consider infinitesimal transition rates $\breve{\Upsilon}_i(x)$, $i = 1,2,...,10$, defined as for $\Upsilon_i(x)$, $i = 1,2,...,10$, with all the parameters having the same values except that μ' is replaced by $\breve{\mu}'$, with $\mu' \geq \breve{\mu}'$. While we have not been able to see how to exploit Theorems S.2 and 3.4 from [13] for these exact rates, we have been able to do this for closely related rates. If we introduce a small approximation in the transition rates of X^ε, namely, $\frac{x_3-1}{2} \approx x_3$ and $\frac{x_4-1}{2} \approx x_4$ in $f_{R121}(x)$ and $f_{R122}(x)$, respectively, then Theorems S.2 and 3.4 in [13] apply with

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix},$$

and $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \ x \preccurlyeq_A w\}$ (see SI - Section S.11). This approximation can be justified by introducing the reasonable assumption that each nucleosome characterized by a repressive modification ($\mathrm{D}^{\mathrm{R}}_1$ and $\mathrm{D}^{\mathrm{R}}_2$) has the ability to catalyze the establishment of the opposite repressive mark on itself. With this approximation, since $a = (0, \mathrm{D}_{\mathrm{tot}}, 0, 0)^T \preccurlyeq_A r = (\mathrm{D}_{\mathrm{tot}}, 0, 0, 0)^T$, then $\breve{h}_{a,r}(\varepsilon) \leq h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon) \leq \breve{h}_{r,a}(\varepsilon)$. Thus, the time to memory loss of the active state increases with higher values of μ', while the time to memory loss of the repressed state decreases with higher values of μ'.

6 Conclusion

In this paper, we provided a mathematical formulation and rigorous proofs to validate the computational findings in [10], showing how the time scale separation between establishment and erasure processes of chromatin modifications affects epigenetic cell memory. To this end, we developed and adapted theory for singularly perturbed continuous time Markov chains and we analyzed the behavior of stationary distributions and mean first passage times as functions of the singular perturbation parameter ε.

We first showed that $\pi(\varepsilon)$ can be expressed as a series expansion (Section 3.1) for sufficiently small ε. We then proved that the limit $\pi(0) = \lim_{\varepsilon \to 0} \pi(\varepsilon)$ is unique and we determined an expression for it (Section 4.1.1). We also provided an iterative procedure for computing all of the higher order terms in the expansion of $\pi(\varepsilon)$ (SI -Section S.2.1). Similarly, for the mean first passage time (MFPT) between states, we first showed there is a Laurent series expansion for sufficiently small ε (Section 3.1, Eq. (3.4)). We then developed a graph based algorithm to identify the order of the leading term in the series expansion (Section 4.2.1), and we also determined the leading coefficient there (Section 4.2.2).

We then applied these tools to the chromatin modification circuit models proposed in [10], to provide a rigorous basis for the computational findings given there (Sections 2, 4, and 5).

Our rigorous derivations of the analytical expressions for the stationary distributions and time to memory loss, and our results on monotonic dependence on parameters, lead to a mechanistic understanding of how ε, μ and μ' affect the stochastic behavior of chromatin modification systems. As an example, our results suggest that higher values of μ and μ' shift mass of the stationary distribution more towards the active state (Sections 5.1.1 and 5.2.1). This finding is consistent with recent experimental results demonstrating that transfection of the DNA methylation eraser enzyme TET1 (represented in our model by higher μ' [10]) into Chinese hamster ovary (CHO-K1) repressed cells causes them to shift towards the active state [27]. More generally, the mechanistic understanding of how ε, μ, and μ' affect the stochastic behavior of chromatin modification systems, as derived in our study, is crucial for determining experimental interventions on molecular players, such as chromatin modifier enzymes, to modulate cell memory. This mechanistic insight is expected to be extremely valuable for applications such as cell fate reprogramming and engineering approaches to cell therapy. Furthermore, the mathematical results and theoretical tools developed in this paper can be applied beyond the scope of the epigenetic cell memory models analyzed in this research work. In fact, they can be applied to all stochastic models that respect the assumptions considered. Future work will investigate how to generalize these results by removing some of these assumptions, including allowing the Markov chain to have countably many states and $Q(0)$ to have ergodic classes as well as absorbing states. While there is some theory for countably many states, such as in [3], the continuous time Markov chains for further applications that we have in mind are not uniformizable and have many transient states for $Q(0)$, and the theory in [3] needs to be generalized for them.

Supplementary information (SI) file: file containing the proofs of the theoretical tools developed in this paper, and detailed mathematical derivations for some of the chromatin modification circuit models analyzed.

Ethics declarations: The authors declare that they have no conflicts of interest.

Data availability: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

REFERENCES

[1] Allis, C. D., Caparros, M.-L., Jenuwein, T. and Reinberg, D. (2015). *Epigenetics, 2nd Edition.* Cold Spring Harbor Laboratory Press.

[2] Anderson, D.F. and Kurtz, T.G. (2011). *Continuous Time Markov Chain Models for Chemical Reaction Networks.* Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology, pp. 3-42.

[3] Altman, E. and Avrachenkov, K.E. and Núñez-Queija, R. (2004). *Perturbation analysis for denumerable markov chains with application to queueing models* Advances in Applied Probability, vol. 36(3): 839–853.

[4] Anderson, D.F. and Kurtz, T.G. (2015). *Stochastic Analysis of Biochemical Systems.* Springer International Publishing.

[5] Avrachenkov K.E. and Haviv, M. (2004). *The first laurent series coefficients for singularly perturbed stochastic matrices.* Linear Algebra and Its Applications, vol. 386, pp. 243–259.

[6] Avrachenkov, K.E., Filar, J.A. and Howlett, P.G. (2013). *Analytic Perturbation Theory and Its*

Applications. SIAM, Philadelphia.

[7] Beltrán J., Landim C. (2010). *Tunneling and Metastability of Continuous Time Markov Chains* Journal of Statistical Physics, vol. 140(6): 1065–1114.

[8] Beltrán J., Landim C. (2012). *Tunneling and metastability of continuous time Markov Chains II, the nonreversible case* Journal of Statistical Physics, vol. 149(4): 598–618.

[9] Blumenthal, R.M. and Getoor, R.K. (1968). *Markov Processes and Potential Theory, 2nd edition.* Academic Press, New York and London.

[10] Bruno, S., Williams, R.J. and Del Vecchio, D. (2022). Epigenetic cell memory: The gene's inner chromatin modification circuit. *PLOS Comp Biol*, Public Library of Science, vol. 18(4): 1 - 27.

[11] Bruno, S., Williams, R.J. and Del Vecchio, D. (2022). Model reduction and stochastic analysis of the histone modification circuit. *2022 European Control Conference (ECC)*, 264-271.

[12] Bruno, S., Williams, R.J. and Del Vecchio, D. (2022). MAnalytical and computational study of the stochastic behavior of a chromatin modification circuit. *2022 IEEE 61st Conference on Decision and Control (CDC)*, 3871-3877.

[13] Campos, F.A., Bruno, S., Fu, Y., Del Vecchio, D. and Williams, R.J. (2023). Comparison theorems for stochastic chemical reaction networks. *Bull Math Biol*, vol. 85(39).

[14] Carey, N. (2013). *The epigenetic revolution.* Columbia University Press.

[15] Cooper, G.M. (2000). *The Cell: A Molecular Approach.* Sunderland (MA): Sinauer Associates.

[16] Del Vecchio, D. and Murray, R.M. (2014). *Biomolecular Feedback Systems.* Princeton University Press.

[17] Dodd, I.B. and Micheelsen, M.A. and Sneppen, K. and Thon, G. (2007). Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. *Cell*, vol. 129.

[18] Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature*, vol. 421.

[19] Hanna, J. and Saha, K. and Pando, B. and van Zon, J. and Lengner, C.J. and Creyghton, M.P. and van Oudenaarden, A. and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature.* 462.

[20] Hassin, R. and Haviv, M. (1992). Mean passage times and nearly uncoupled Markov chains. *SIAM Journal on Discrete Mathematics.* 5 (3), 386-397.

[21] Heard, E. and Martienssen, R.A. (2014). Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell*, vol. 157.

[22] Huang, S. and Litt, M. and Blakey, C. A. (2015). *Epigenetic Gene Expression and Regulation.* Academic Press.

[23] Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* New York, Springer-Verlag. Graduate Texts in Mathematics 113.

[24] Kim, M. and Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med*, vol. 49, no. e322.

[25] Meyer, Jr. and Carl, D. (1975). The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, vol. 17, no. 3, pp. 443–464.

[26] Norris, J.R. (1997). *Markov Chains.* Cambridge University Press.

[27] Palacios, S. and Bruno, S. and Weiss, R. and Salibi, E. and Kane, A. and Ilia, K. and Del Vecchio, D. (2024). Analog epigenetic cell memory by graded DNA methylation. *bioRxiv: https://doi.org/10.1101/2024.02.13.580200* .

[28] Pan, Z.G. and Basar, T. (1999). H infintity control of large-scale jump linear systems via averaging and aggregation. *International Journal of Control*, vol. 72, no. 10, pp. 866–881.

[29] Phillips, R.G. and Kokotovic, P.V (1981). A singular perturbation approach to modelling and control of Markov chains. *IEEE Trans. Automat. Control*, vol. 26, pp. 1087–1094.

[30] Ptashne, M. (2013). Epigenetics: Core misconcept. *Proc. Natl. Acad. Sci*, vol. 110.

[31] Xiong, W. and Ferrell Jr, J.E. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, vol. 426.

[32] Yin, G.G. and Zhang, Q. (2013). *Continuous-Time Markov Chains and Applications. A Two-Time-Scale Approach, Second Edition.* Springer, New York, NY.

[33] Zhang, T. and Cooper, S. and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. *EMBO Reports*, vol. 16.

Analysis of singularly perturbed stochastic chemical reaction networks motivated by applications to epigenetic cell memory

Simone Bruno[1,*], Felipe A. Campos[2,*], Yi Fu[2], Domitilla Del Vecchio[1], and Ruth J. Williams[2]

[1] *Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Emails:* (sbruno,ddv)@mit.edu

[2] *Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0112. Email:* (fcamposv,yif064,rjwilliams)@ucsd.edu

[*] *These authors contributed equally: S. Bruno and F. A. Campos*

Supplementary Information (SI)

S.1 Some results in probability Let $\mathcal{X}$ be a finite set. Recall the notation for matrices introduced in Section 1.5.

Lemma S.1. *Let $X = \{X(t) : t \geq 0\}$ be a continuous time Markov chain with state space $\mathcal{X}$ and infinitesimal generator $Q = (Q_{x,y})_{x,y\in\mathcal{X}}$. Then, the number of recurrent classes for X is equal to* $\text{nullity}(Q^T) = \text{nullity}(Q)$.

Proof. Since Q is a square matrix, the Rank plus Nullity Theorem yields that $\text{nullity}(Q^T) = \text{nullity}(Q)$. Now, consider $\lambda > \max_{x\in\mathcal{X}} |Q_{x,x}|$ and define $P := I + Q/\lambda$, where I is the identity matrix of size $|\mathcal{X}| \times |\mathcal{X}|$. The matrix P is stochastic and such that for every $x \neq y$ in $\mathcal{X}$, $P_{x,y} > 0$ if and only if $Q_{x,y} > 0$. As a consequence, the recurrent classes of X are the same as the recurrent classes of P. By Theorem IV.2.4 in Isaacson and Madsen [11], the number of recurrent classes of P is equal to the maximum number of linearly independent left eigenvectors satisfying $\pi P = \pi$. By observing that $\pi P = \pi$ if and only if $\pi Q = 0$, we see that this latter quantity is equal to $\text{nullity}(Q^T)$. ■

The following is Proposition 6.3 in Asmussen [1].

Proposition S.2. *Let $(P_{x,y})_{x,y\in\mathcal{X}}$ be a nonnegative substochastic matrix ($P\mathbb{1} \leq \mathbb{1}$) such that for each $x \in \mathcal{X}$ there are $z_1, \ldots, z_m, y \in \mathcal{X}$ such that $P_{x,z_1} P_{z_1,z_2} \ldots P_{z_m,y} > 0$ and $\sum_{z\in\mathcal{X}} P_{y,z} < 1$. Then,* $\text{spr}(P) < 1$.

We use Proposition S.2 in order to obtain invertibility for some matrices, as in the next result.

Lemma S.3. *Let $X = \{X(t) : t \geq 0\}$ be an irreducible continuous time Markov chain with state space $\mathcal{X}$ and with an embedded discrete time Markov chain with transition matrix P. Consider a nonempty set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{B} \neq \mathcal{X}$ and consider $P^{\mathcal{B}^c}$ to be the matrix obtained by removing the columns and rows of P corresponding to states in $\mathcal{B}$. Then, $I - P^{\mathcal{B}^c}$ is invertible and its inverse is given by the absolutely convergent series $\sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k$, where $(P^{\mathcal{B}^c})^0 = I$.*

Proof. Observe that $P^{\mathcal{B}^c} = (P_{x,y})_{x,y\in\mathcal{B}^c}$ is a nonnegative substochastic matrix. Since X is an irreducible continuous time Markov chain, its embedded discrete time Markov chain is also irreducible. Thus, for each $x \in \mathcal{B}^c$, there exist $z_1, \ldots, z_m, y \in \mathcal{B}^c$ and $\tilde{y} \in \mathcal{B}$ such that $P_{x,z_1} P_{z_1,z_2} \ldots P_{z_m,y} P_{y,\tilde{y}} > 0$. Then, $P^{\mathcal{B}^c}_{x,z_1} P^{\mathcal{B}^c}_{z_1,z_2} \ldots P^{\mathcal{B}^c}_{z_m,y} > 0$ and $\sum_{z\in\mathcal{B}^c} P^{\mathcal{B}^c}_{y,z} = \sum_{z\in\mathcal{B}^c} P_{y,z} < 1$ since $P_{y,\tilde{y}} > 0$ and $\sum_{z\in\mathcal{X}} P_{y,z} = 1$. By Proposition S.2, $\text{spr}(P^{\mathcal{B}^c}) < 1$. This fact, together with Theorem 5.6.15 in Horn & Johnson [10] yields the convergence of $\sum_{k=0}^{\infty} (P^{\mathcal{B}^c})^k$. Moreover,

$(I - P^{\mathcal{B}^c})\sum_{k=0}^{\infty}(P^{\mathcal{B}^c})^k = \sum_{k=0}^{\infty}(P^{\mathcal{B}^c})^k(I - P^{\mathcal{B}^c}) = I$, which yields the desired result. ∎

We will use the following continuous time analogue of Proposition S.2.

Lemma S.4. *Let $(Q_{x,y})_{x,y\in\mathcal{X}}$ be a matrix such that $Q\mathbb{1} \le 0$ and such that $Q_{x,x} \le 0$ for each $x \in \mathcal{X}$ and $Q_{x,y} \ge 0$ for each $x \ne y \in \mathcal{X}$. In addition, suppose that for each $x \in \mathcal{X}$ there are distinct $z_1, \ldots, z_m, y \in \mathcal{X}$ different from x such that $Q_{x,z_1}Q_{z_1,z_2}\ldots Q_{z_m,y} > 0$ and $\sum_{z\in\mathcal{X}} Q_{y,z} < 0$. Then, for every $\upsilon \in \mathrm{sp}(Q)$, the real part of υ is negative. In particular, Q is invertible.*

Proof. Consider $\lambda > \max_{x\in\mathcal{X}} |Q_{x,x}|$ and define $P := I + Q/\lambda$, where I is the identity matrix of size $|\mathcal{X}| \times |\mathcal{X}|$. The matrix P is nonnegative, substochastic and such that $P_{x,y} = \frac{1}{\lambda}Q_{x,y}$ for every $x \ne y \in \mathcal{X}$. With these elements, we obtain that for each $x \in \mathcal{X}$ there are distinct $z_1, \ldots, z_m, y \in \mathcal{X}$ such that $P_{x,z_1}P_{z_1,z_2}\ldots P_{z_m,y} > 0$ and where $\sum_{z\in\mathcal{X}} P_{y,z} = 1 + \frac{1}{\lambda}\sum_{z\in\mathcal{X}} Q_{y,z} < 1$. Proposition S.2 yields that $\mathrm{spr}(P) < 1$. By observing that $\upsilon \in \mathrm{sp}(Q)$ implies that $1 + \frac{\upsilon}{\lambda} \in \mathrm{sp}(P)$, we obtain that $1 > |1 + \frac{\upsilon}{\lambda}| \ge |1 + \frac{\Re(\upsilon)}{\lambda}|$ where $\Re(\upsilon)$ is the real part of υ. The latter inequality implies that $\Re(\upsilon) < 0$. ∎

Consider a nonempty set $\mathcal{B} \subseteq \mathcal{X}$ such that $\mathcal{B} \ne \mathcal{X}$ and a Q-matrix written as

$$Q = \begin{array}{c} \\ \mathcal{B} \\ \mathcal{B}^c \end{array} \begin{array}{c} \begin{array}{cc} \mathcal{B} & \mathcal{B}^c \end{array} \\ \left(\begin{array}{c|c} Q^{\mathcal{B}} & S \\ R & Q^{\mathcal{B}^c} \end{array}\right) \end{array}. \tag{S.1}$$

Consider a process $X = \{X(t) : t \ge 0\}$ defined on a measurable space $(\Omega, \mathcal{F})$ and a collection of probability measures $\{\mathbb{P}_x : x \in \mathcal{X}\}$ on $(\Omega, \mathcal{F})$ such that for every $x \in \mathcal{X}$, X is a continuous time Markov chain under $\mathbb{P}_x$ with infinitesimal generator given by Q and such that $\mathbb{P}_x[X(0) = x] = 1$. Consider the stopping time $\tau_{\mathcal{B}} := \inf\{t \ge 0 : X(t) \in \mathcal{B}\}$.

Lemma S.5. *The following are equivalent:*

(i) For every $x \in \mathcal{B}^c$, there exists a $z \in \mathcal{B}$ such that x leads to z under Q, i.e., there are distinct $x_1, ..., x_m \in \mathcal{B}^c$, different from x, such that $Q_{x,x_1}, Q_{x_1,x_2}, ..., Q_{x_m,z} > 0$.

(ii) $Q^{\mathcal{B}^c}$ is invertible.

(iii) $\mathbb{E}_x[\tau_{\mathcal{B}}] < \infty$ for every $x \in \mathcal{B}^c$.

If any of $(i) - (iii)$ hold, then

$$\mathbb{P}_x\left[X(\tau_{\mathcal{B}}) = y\right] = (-(Q^{\mathcal{B}^c})^{-1}R)_{x,y} \tag{S.2}$$

for every $x \in \mathcal{B}^c$ and $y \in \mathcal{B}$. Moreover, if $\mathcal{B}^c$ is a set of transient states for X, then $(i) - (iii)$ hold.

Part of the results in Lemma S.5 appear as Lemma 1 in Gaver et al. [8] for the case where $Q^{\mathcal{B}}$ and S are the zero matrix. For completeness, we provide a self-contained proof here.

Proof. The implication $(i) \Rightarrow (ii)$ is a consequence of Lemma S.4 with $\mathcal{B}^c, Q^{\mathcal{B}^c}$ in place of $\mathcal{X}, Q$ there.

In the following, recall that for every $x \in \mathcal{X}$ and function $f : \mathcal{X} \longrightarrow \mathbb{R}$, the process

$$M_f^{\mathcal{B}}(t) := f(X(t \wedge \tau_{\mathcal{B}})) - f(X(0)) - \int_0^{t\wedge\tau_{\mathcal{B}}} \mathcal{L}f(X(s))ds, \qquad t \ge 0, \tag{S.3}$$

is a martingale under $\mathbb{P}_x$ (see Theorem 3.32 in [15]), where $\mathcal{L}f(y) := \sum_{z\in\mathcal{X}} Q_{yz}f(z)$ for $y \in \mathcal{X}$.

For $(ii) \Rightarrow (iii)$, consider the function $f(y) := -[(Q^{\mathcal{B}^c})^{-1}\mathbb{1}]_y \mathbb{1}_{\mathcal{B}^c}(y)$ for $y \in \mathcal{X}$. The reader may verify that $\mathcal{L}f(y) = -1$ for every $y \in \mathcal{B}^c$. Therefore, for an $x \in \mathcal{B}^c$, taking expectations in (S.3) yields $f(x) - \mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = \mathbb{E}_x[t \wedge \tau_{\mathcal{B}}]$ for every $t \geq 0$. Hence, $\mathbb{E}_x[t \wedge \tau_{\mathcal{B}}] \leq 2\sup_{x\in\mathcal{X}} |f(x)|$ for every $t \geq 0$ and we conclude the desired result by letting $t \to \infty$.

For $(iii) \Rightarrow (i)$, we prove that not (i) implies not (iii). Suppose there exists $x \in \mathcal{B}^c$ such that no point of $\mathcal{B}$ is accessible from x. Then, $\tau_{\mathcal{B}} = \infty$ $\mathbb{P}_x$-a.s., so (iii) does not hold.

For (S.2), consider $x \in \mathcal{B}^c$, $y \in \mathcal{B}$ and the martingale $M_f^{\mathcal{B}}$ with $f(z) = (-(Q^{\mathcal{B}^c})^{-1}R)_{z,y}\mathbb{1}_{\mathcal{B}^c}(z)$ $+\mathbb{1}_{\{y\}}(z)$ for $z \in \mathcal{X}$. The reader may verify that $\mathcal{L}f(x) = 0$ for $x \in \mathcal{B}^c$, which yields that $\mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = f(x)$ for every $t \geq 0$. If any of $(i) - (iii)$ hold, then $\tau_{\mathcal{B}} < \infty$, $\mathbb{P}_x$-a.s., and on letting $t \to \infty$ and using bounded convergence, we obtain $\mathbb{E}_x[f(X(\tau_{\mathcal{B}}))] = f(x)$, which implies $\mathbb{P}_x[X(\tau_{\mathcal{B}}) = y] = (-(Q^{\mathcal{B}^c})^{-1}R)_{x,y}$.

Now, suppose that every $x \in \mathcal{B}^c$ is transient. Then, $\mathbb{P}_x(\tau_{\mathcal{B}} < \infty) = 1$ for each $x \in \mathcal{B}^c$. For a proof by contradiction, suppose that $Q^{\mathcal{B}^c}$ is not invertible, which implies the existence of a nonzero vector $v = (v(x))_{x\in\mathcal{B}^c} \neq 0$ such that $Q^{\mathcal{B}^c}v = 0$. Then, consider the function $f(y) = v(y)\mathbb{1}_{\mathcal{B}^c}(y)$, for which $\mathcal{L}f(y) = 0$ for $y \in \mathcal{B}^c$. Consider an $x \in \mathcal{B}^c$ such that $v(x) \neq 0$, then $M_f^{\mathcal{B}}(t) = f(X(t \wedge \tau_{\mathcal{B}})) - v(x)$ is a bounded martingale. On taking expectations we have $\mathbb{E}_x[f(X(t \wedge \tau_{\mathcal{B}}))] = v(x)$. Since the states in $\mathcal{B}^c$ are transient, $X(\cdot)$ will $\mathbb{P}_x$-a.s. leave $\mathcal{B}^c$. Then letting $t \to \infty$ and using bounded convergence yields $0 = v(x)$ which is the desired contradiction. Hence (ii) (as well as $(i), (iii)$) must hold. ■

Lemma S.5 has a useful consequence in terms of occupations times. In the above context, consider the occupation time of $\mathcal{B}$ by the Markov chain X up to time $t \geq 0$: $\chi_{\mathcal{B}}(t) = \int_0^t \mathbb{1}_{\mathcal{B}}(X(s))ds$. Denote by $\chi_{\mathcal{B}}(\infty) = \lim_{t\to\infty} \chi_{\mathcal{B}}(t) = \int_0^\infty \mathbb{1}_{\mathcal{B}}(X(s))ds$.

Lemma S.6. *Suppose that*

$$\mathbb{P}_y[\tau_{\mathcal{B}} < \infty] = 1 \text{ for all } y \in \mathcal{B}^c. \tag{S.4}$$

Then $\mathbb{P}_x[\chi_{\mathcal{B}}(\infty) = \infty] = 1$ *for every* $x \in \mathcal{X}$.

Remark S.7. If any of the conditions (i)-(iii) in Lemma S.5 holds, then (S.4) holds.

Proof. Fix $x \in \mathcal{X}$. Let $\sigma_{-1} = 0$ and $\sigma_0 = \inf\{t \geq \sigma_{-1} : X(t) \in \mathcal{B}\}$. Then, inductively define for $k = 0, 1, 2, \ldots$, $\sigma_{2k+1} = \inf\{t \geq \sigma_{2k} : X(t) \in \mathcal{B}^c\}$ and $\sigma_{2k+2} = \inf\{t \geq \sigma_{2k+1} : X(t) \in \mathcal{B}\}$. Using (S.4) and the strong Markov property of X, we have

$$\sigma_{2k} < \infty \ \mathbb{P}_x\text{-a.s. on } \{\sigma_{2k-1} < \infty\} \tag{S.5}$$

for $k = 0, 1, 2, \ldots$, and

$$\chi_{\mathcal{B}}(\infty) = \sum_{k=0}^{\infty} \mathbb{1}_{\{\sigma_{2k}<\infty\}}(\sigma_{2k+1} - \sigma_{2k}) = \lim_{N\to\infty} \sum_{k=0}^{N} \mathbb{1}_{\{\sigma_{2k}<\infty\}}(\sigma_{2k+1} - \sigma_{2k}),$$

where terms in the sum indexed by $k : \sigma_{2k} = \infty$ are taken to be zero. Now, $\mathbb{P}_x$-a.s.,

$$\prod_{k=0}^{N} \exp(-\mathbb{1}_{\{\sigma_{2k}<\infty\}}(\sigma_{2k+1} - \sigma_{2k})) = \prod_{k=0}^{N} \mathbb{1}_{\{\sigma_{2k}<\infty\}} \exp(-(\sigma_{2k+1} - \sigma_{2k})),$$

where we used (S.5) and the fact that $\sigma_{-1} = 0$, to conclude that the product is zero $\mathbb{P}_x$-a.s., if $\{\sigma_{2k} = \infty\}$ for any $k \in \{0, 1, \dots, N\}$. Hence,

$$\begin{aligned}
&\mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] \\
&= \lim_{N\to\infty} \mathbb{E}_x\left[\prod_{k=0}^{N} \mathbb{1}_{\{\sigma_{2k}<\infty\}} \exp(-(\sigma_{2k+1}-\sigma_{2k}))\right] \\
&= \lim_{N\to\infty} \mathbb{E}_x\left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k}<\infty\}} \exp(-(\sigma_{2k+1}-\sigma_{2k}))\mathbb{1}_{\{\sigma_{2N}<\infty\}}\mathbb{E}_x\left[\exp(-(\sigma_{2N+1}-\sigma_{2N}))|X(\sigma_{2N})\right]\right].
\end{aligned}$$

On $\{\sigma_{2N} < \infty\}$, we have $X(\sigma_{2N}) \in \mathcal{B}$ and $X(t) \in \mathcal{B}$ for $\sigma_{2N} \le t < \sigma_{2N+1}$. Hence, for $a > \max_{y\in\mathcal{B}} |Q_{y,y}|$, using the strong Markov property, on $\{\sigma_{2N} < \infty\}$, conditioned on $X(\sigma_{2N})$, $\sigma_{2N+1} - \sigma_{2N}$ stochastically dominates an exponential random variable with parameter a and so

$$\mathbb{E}_x\left[\exp(-(\sigma_{2N+1}-\sigma_{2N}))|X(\sigma_{2N})\right] \le \int_0^\infty e^{-t} a e^{-at} dt = \frac{a}{1+a}.$$

Similarly, for $k = N-1, \dots, 0$, on $\{\sigma_{2k} < \infty\}$,

$$\mathbb{E}_x\left[\exp(-(\sigma_{2k+1}-\sigma_{2k}))|X(\sigma_{2k})\right] \le \frac{a}{1+a}. \tag{S.6}$$

Then,

$$\begin{aligned}
\mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] &\le \limsup_{N\to\infty} \mathbb{E}_x\left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k}<\infty\}} \exp(-(\sigma_{2k+1}-\sigma_{2k}))\mathbb{1}_{\{\sigma_{2N}<\infty\}}\frac{a}{1+a}\right] \\
&\le \limsup_{N\to\infty} \mathbb{E}_x\left[\prod_{k=0}^{N-1} \mathbb{1}_{\{\sigma_{2k}<\infty\}} \exp(-(\sigma_{2k+1}-\sigma_{2k}))\frac{a}{1+a}\right].
\end{aligned}$$

Repeatedly conditioning on $\{\sigma_{2k} < \infty\}$, for $k = N-1, \dots, 0$ and using (S.6), we obtain

$$\mathbb{E}_x[\exp(-\chi_{\mathcal{B}}(\infty))] \le \limsup_{N\to\infty} \left(\frac{a}{1+a}\right)^{N+1} = 0.$$

Hence, $\mathbb{P}_x[\chi_{\mathcal{B}}(\infty) = \infty] = 1$. ■

Lemma S.8. *Let $X = \{X(t) : t \ge 0\}$ be a continuous time Markov chain with state space $\mathcal{X}$ and infinitesimal generator $Q = (Q_{x,y})_{x,y\in\mathcal{X}}$. Suppose there is an absorbing state $y \in \mathcal{X}$. If there are distinct states $z_1, \dots, z_m \in \mathcal{X}$ different from x and y such that $Q_{x,z_1}Q_{z_1,z_2}\dots Q_{z_m,y} > 0$, then x is a transient state for X.*

Proof. Since $Q_{x,z_1}Q_{z_1,z_2}\dots Q_{z_m,y} > 0$, we have $\mathbb{P}_x[X(t_0) = y] > 0$ for some $t_0 > 0$ (see Theorem 3.2.1 in [16]). Thus $\mathbb{P}_x[X(t) \ne x$ for all $t \ge t_0] \ge \mathbb{P}_x[X(t_0) = y]\mathbb{P}[X(t) = y$ for all $t > t_0 | X(t_0) = y] = \mathbb{P}_x[X(t_0) = y] > 0$, which means that x is a transient state. ■

S.2 Additional results for stationary distributions

S.2.1 Higher order terms for linear perturbations Under the assumption of the perturbation being linear (which is the case for our chromatin modification circuit models), we now provide an iterative procedure for computing all of the terms in the series expansion of $\pi(\cdot)$. Additional results for characterizing some of these terms are given in SI - Sections S.2.3, S.2.4.

Theorem S.9. *Suppose Assumptions 4.1, 4.2 and 4.5 hold. Then, the following hold for the sequence* $\{\pi^{(k)} : k \geq 0\}$ *in* (3.7)*:*

(i) $\pi^{(0)} = [\alpha^{(0)}, \beta^{(0)}] = [\alpha, 0]$ *where* α *is the unique probability vector on* $\mathcal{A}$ *such that* $\alpha Q_{\mathcal{A}} = 0$,

(ii) *for every* $k \geq 1$, $\pi^{(k)} = [\alpha^{(k)}, \beta^{(k)}]$, *where*

$$\beta^{(k)} = (\alpha^{(k-1)} S_1 + \beta^{(k-1)} T_1)(-T_0)^{-1} \tag{S.7}$$

and $\alpha^{(k)}$ *is the unique vector such that*

$$\alpha^{(k)} Q_{\mathcal{A}} = -\beta^{(k)}[R_1 + T_1(-T_0)^{-1} R_0], \tag{S.8}$$

$$\alpha^{(k)} \mathbb{1} = -\beta^{(k)} \mathbb{1}. \tag{S.9}$$

Moreover, if $|\mathcal{A}| \geq 2$, *for every* $k \geq 1$ *we obtain*

$$\alpha^{(k)} = \tilde{\alpha}^{(k)} + (-\beta^{(k)} \mathbb{1} - \tilde{\alpha}^{(k)} \mathbb{1})\alpha, \tag{S.10}$$

where $\tilde{\alpha}^{(k)} := -\beta^{(k)}(R_1 + T_1(-T_0)^{-1} R_0) Q_{\mathcal{A}}^{\dagger}$ *for* $k \geq 1$ *and* $Q_{\mathcal{A}}^{\dagger}$ *is a generalized inverse of* $Q_{\mathcal{A}}$ [4].

The proof of Theorem S.9 is given in SI - Section S.2.2.

S.2.2 Proofs of Lemmas 4.1, 4.3, and 4.4 and Theorem S.9

Proof of Lemma 4.1

Proof. It has already been established before Lemma 4.1 that $\pi(0) = [\alpha, 0]$. By equating to zero the coefficients of the terms ε^m for $m = 0, 1, \ldots$ in the series $(\sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)})(\sum_{k=0}^{\infty} \varepsilon^k Q^{(k)})$, we obtain that $\sum_{k=0}^{m} \pi^{(k)} Q^{(m-k)} = 0$ for every $m \geq 0$. In particular, $\pi^{(1)} Q^{(0)} + \pi^{(0)} Q^{(1)} = 0$, which yields,

$$[\alpha^{(1)}, \beta^{(1)}] \left(\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array} \right) + [\alpha, 0] \left(\begin{array}{c|c} A_1 & S_1 \\ \hline R_1 & T_1 \end{array} \right) = 0.$$

From this, we obtain two equations:

$$\beta^{(1)} R_0 + \alpha A_1 = 0 \tag{S.11}$$

[4] A generalized inverse $Q_{\mathcal{A}}^{\dagger}$ of $Q_{\mathcal{A}}$ is such that $Q_{\mathcal{A}} Q_{\mathcal{A}}^{\dagger} Q_{\mathcal{A}} = Q_{\mathcal{A}}$. The Moore-Penrose inverse is a generalized inverse. The deviation matrix for $\hat{X}_{\mathcal{A}}$ is $D = (-Q_{\mathcal{A}} + \mathbb{1}\alpha)^{-1} - \mathbb{1}\alpha$, and $-D$ is also a generalized inverse of $Q_{\mathcal{A}}$. Meyer [4] suggested that $-D$ is a better generalized inverse to use than the Moore-Penrose inverse since it can be computed efficiently and embeds answers concerning the transitory behavior of the Markov chain. Avrachenkov et al. [2], in the context of discrete time Markov chains, use a suitable deviation matrix when they need a generalized inverse. Here, if we take $Q_{\mathcal{A}}^{\dagger} = -D$, then the term $\tilde{\alpha}^{(k)} \mathbb{1}$ in (S.10) is equal to zero since $D\mathbb{1} = 0$ and then $\alpha^{(k)} = \beta^{(k)}((R_1 + T_1(-T_0)^{-1} R_0)D - \mathbb{1}\alpha)$.

and

$$\beta^{(1)} T_0 + \alpha S_1 = 0. \tag{S.12}$$

Since T_0 is invertible, from (S.12) we obtain (4.5). We conclude by substituting this formula for $\beta^{(1)}$ in (S.11). ■

Proof of Lemma 4.3

Proof. By following the proof of Lemma 2 in Gaver, Jacobs & Latouche [8] and using formula (S.2), we can prove that the transition rates between $x \neq y \in \mathcal{A}$ for $\hat{X}_{\mathcal{A}}$ are given by $(Q_{\mathcal{A}})_{x,y}$. In essence, the argument is as follows. From the state $x \in \mathcal{A}$, the Markov chain $\tilde{X}$ may move to $y \in \mathcal{A}$ in two ways that lead to a one-step transition for $\hat{X}_{\mathcal{A}}$. First, it could happen that $\tilde{X}$ jumps directly to y at a rate of $(A_1)_{x,y}$. Second, the chain $\tilde{X}$ may go to some state $z \in \mathcal{T}$ at a rate $(S_1)_{x,z}$ and from there, jump between states in $\mathcal{T}$ until getting back to $\mathcal{A}$ at the state $y \in \mathcal{A}$. By (S.2), this happens with probability $((-T_0)^{-1} R_0)_{z,y}$. Putting this all together, the rate of transition for $\hat{X}_{\mathcal{A}}$ from x to y will be

$$(A_1)_{x,y} + \sum_{z \in \mathcal{T}} (S_1)_{x,z} ((-T_0)^{-1} R_0)_{z,y} = (Q_{\mathcal{A}})_{x,y}. \tag{S.13}$$ ■

Proof of Lemma 4.4

Proof. Consider $x \neq y \in \mathcal{A}$. Then, there exists a sequence of states $x_0 = x, x_1, \ldots, x_m = y$ in $\mathcal{C}$ such that $\tilde{Q}_{x,x_1} \tilde{Q}_{x_1,x_2} \ldots \tilde{Q}_{x_{m-1},y} > 0$. Roughly speaking, the proof follows by erasing the times that $\tilde{X}$ is outside of $\mathcal{A}$. We now give the details. Consider $i \in \{0, 1, \ldots, m-1\}$ with $x_i \in \mathcal{A}$. If $x_{i+1} \in \mathcal{A}$, then, by (S.13), $(Q_{\mathcal{A}})_{x_i,x_{i+1}} \geq (A_1)_{x_i,x_{i+1}} = \tilde{Q}_{x_i,x_{i+1}} > 0$. If $x_{i+1} \in \mathcal{T}$, consider the path of states $x_i, x_{i+1}, \ldots, x_k$ for $0 \leq i < k \leq m$ such that $x_i, x_k \in \mathcal{A}$ and $x_{i+1}, \ldots, x_{k-1} \in \mathcal{T}$. Since x_{i+1} leads to x_k for $\tilde{X}$, then $\mathbb{P}_{x_{i+1}}[\tilde{X}(\tau_{\mathcal{A}}) = x_k] > 0$ where $\tau_{\mathcal{A}} := \inf\{t \geq 0 : \tilde{X}(t) \in \mathcal{A}\}$. By (S.2), this means that $((-T_0)^{-1} R_0)_{x_{i+1},x_k} > 0$ which yields $(Q_{\mathcal{A}})_{x_i,x_k} \geq (S_1)_{x_i,x_{i+1}} ((-T_0)^{-1} R_0)_{x_{i+1},x_k} = \tilde{Q}_{x_i,x_{i+1}} ((-T_0)^{-1} R_0)_{x_{i+1},x_k} > 0$. These observations yield a sequence of states $x_0 = x, x_{i_1}, \ldots, x_{i_j} = y$ in $\mathcal{A}$ such that $(Q_{\mathcal{A}})_{x,x_{i_1}} (Q_{\mathcal{A}})_{x_{i_1},x_{i_2}} \ldots (Q_{\mathcal{A}})_{x_{i_{j-1}},y} > 0$. ■

Proof of Theorem S.9

Proof. Point (i) was established in Theorem 4.2. For (ii), we equate to zero the coefficients of the terms ε^m for $m = 0, 1, 2, \ldots$ in the terms of the series, $(\sum_{k=0}^{\infty} \varepsilon^k \pi^{(k)})(Q^{(0)} + \varepsilon Q^{(1)})$ to obtain that $\pi^{(0)} Q^{(0)} = 0$ and $\pi^{(k)} Q^{(0)} + \pi^{(k-1)} Q^{(1)} = 0$ for every $k \geq 1$. The latter requires that for all $k \geq 1$,

$$[\alpha^{(k)}, \beta^{(k)}] \left(\begin{array}{c|c} 0 & 0 \\ \hline R_0 & T_0 \end{array} \right) + [\alpha^{(k-1)}, \beta^{(k-1)}] \left(\begin{array}{c|c} A_1 & S_1 \\ \hline R_1 & T_1 \end{array} \right) = 0.$$

Now, this yields two equations:

$$\beta^{(k)} R_0 + \alpha^{(k-1)} A_1 + \beta^{(k-1)} R_1 = 0, \tag{S.14}$$

$$\beta^{(k)} T_0 + \alpha^{(k-1)} S_1 + \beta^{(k-1)} T_1 = 0. \tag{S.15}$$

For $\beta^{(k)}$, we obtain the relation (S.7) directly from (S.15) for all $k \geq 1$. For $\alpha^{(k)}$, let's see first that it satisfies (S.8). From (S.7), for all $k \geq 1$, $\beta^{(k+1)} = (\alpha^{(k)} S_1 + \beta^{(k)} T_1)(-T_0)^{-1}$ and using this in (S.14) (with k replaced by $k+1$) we obtain for all $k \geq 1$

$$(\alpha^{(k)} S_1 + \beta^{(k)} T_1)(-T_0)^{-1} R_0 + \alpha^{(k)} A_1 + \beta^{(k)} R_1 = 0. \tag{S.16}$$

By rearranging (S.16) and using (4.6), we obtain (S.8) for all $k \geq 1$. On the other hand, since $\langle \pi^{(k)}, \mathbb{1} \rangle = 0$ for every $k \geq 1$, we obtain (S.9).

For the uniqueness of $\alpha^{(k)}$, for all $k \geq 1$, if $|\mathcal{A}| = 1$, then $\alpha^{(k)}$ has only one entry and it is determined uniquely by (S.9). If $|\mathcal{A}| \geq 2$, consider another solution $\gamma^{(k)}$ of (S.8) and (S.9), where $(\gamma^{(k)})^T \in \mathbb{R}^{|\mathcal{A}|}$. By Assumption 4.2 and Lemma S.1, $\dim(\ker((Q_{\mathcal{A}})^T)) = 1$ and therefore, by (S.8), $\alpha^{(k)} - \gamma^{(k)} = c\alpha$ for some $c \in \mathbb{R}$. Using (S.9), then $0 = \alpha^{(k)} \mathbb{1} - \gamma^{(k)} \mathbb{1} = c\alpha \mathbb{1} = c$, and therefore $c = 0$, and $\alpha^{(k)} = \gamma^{(k)}$.

For existence of a solution $\alpha^{(k)}$ of (S.8)-(S.9), using the properties

$$R_0 \mathbb{1} + T_0 \mathbb{1} = 0 \quad \text{and} \quad R_1 \mathbb{1} + T_1 \mathbb{1} = 0,$$

we have that

$$\begin{aligned}
(R_1 + T_1(-T_0)^{-1} R_0)\mathbb{1} &= R_1 \mathbb{1} + T_1(-T_0)^{-1} R_0 \mathbb{1} \\
&= -T_1 \mathbb{1} + T_1(-T_0)^{-1} R_0 \mathbb{1} \\
&= T_1(-T_0)^{-1}(T_0 \mathbb{1} + R_0 \mathbb{1}) \\
&= 0.
\end{aligned}$$

Then, since $\dim(\ker(Q_{\mathcal{A}})) = \dim(\ker((Q_{\mathcal{A}})^T)) = 1$ and $\mathbb{1} \in \ker(Q_{\mathcal{A}})$, we have

$$(-\beta^{(k)}(R_1 + T_1(-T_0)^{-1} R_0))^T \in \ker(Q_{\mathcal{A}})^{\perp} = \text{range}((Q_{\mathcal{A}})^T),$$

and so (S.8) has a solution and (S.9) will determine the multiple of α to add to any particular solution to obtain the unique solution $\alpha^{(k)}$ of both equations.

Furthermore, if $Q_{\mathcal{A}}^{\dagger}$ is a generalized inverse of $Q_{\mathcal{A}}$, then

$$\tilde{\alpha}^{(k)} := -\beta^{(k)}(R_1 + T_1(-T_0)^{-1} R_0) Q_{\mathcal{A}}^{\dagger} \tag{S.17}$$

is a solution to (S.8) (see [12] for an exposition). Similar to the uniqueness argument, $\alpha^{(k)} - \tilde{\alpha}^{(k)} = c\alpha$ for some $c \in \mathbb{R}$. By (S.9), $c = -\beta^{(k)} \mathbb{1} - \tilde{\alpha}^{(k)} \mathbb{1}$ and we obtain (S.10). ■

S.2.3 Additional characterization of zeroth and first order terms for linear perturbations via restricted processes

In this section, assume that Assumptions 4.1 and 4.5 hold. We will also sometimes assume Assumptions 4.3 or 4.4 hold. We will explore additional characterizations of α and $\beta^{(1)}$. Under Assumptions 4.1 and 4.5, $A(\varepsilon)$ (defined in (4.2)) corresponds to εA_1 for every $0 \leq \varepsilon < \varepsilon_0$. Since $Q(\varepsilon)$ is irreducible for every $0 < \varepsilon < \varepsilon_0$, from Lemma S.5 (with $\mathcal{A}^c$ in place of $\mathcal{B}$ and $Q(\varepsilon)$ in place of Q there), we obtain that εA_1 is invertible for $0 < \varepsilon < \varepsilon_0$, and therefore A_1 is invertible. This will be an important fact for the coming results.

Consider the matrix $\tilde{Q}$ introduced in (4.7). For a continuous time Markov chain $\tilde{X}$ with infinitesimal generator $\tilde{Q}$, denote by $\chi_{\mathcal{T}}(t)$ the occupation time of $\mathcal{T}$ by the Markov chain $\tilde{X}$

up to time $t \geq 0$, with its associated limit $\chi_{\mathcal{T}}(\infty) = \lim_{t\to\infty} \chi_{\mathcal{T}}(t) = \int_0^\infty \mathbb{1}_{\mathcal{T}}(X(t))dt$. Since A_1 is invertible, by Lemma S.5 (with $\mathcal{B} = \mathcal{T}$ and $Q = \tilde{Q}$) and Lemma S.6 we have that $\mathbb{P}[\chi_{\mathcal{T}}(\infty) = \infty] = 1$.

Consider the process $\hat{X}_{\mathcal{T}}$ as in (4.8), but with $\mathcal{A}$ replaced by $\mathcal{T}$, which corresponds to observing $\tilde{X}$ only on the time intervals where $\tilde{X}$ is in $\mathcal{T}$. The process $\hat{X}_{\mathcal{T}}$ is a continuous time Markov chain on $\mathcal{T}$. Consider the matrix

$$Q_{\mathcal{T}} := T_0 + R_0(-A_1)^{-1}S_1, \tag{S.18}$$

which by Lemma S.15 is a Q-matrix. Similarly to Lemma 4.3, we can show that $Q_{\mathcal{T}}$ is the infinitesimal generator of $\hat{X}_{\mathcal{T}}$. Our previous assumptions relate to $\hat{X}_{\mathcal{T}}$ in the following way.

Lemma S.10. *Suppose Assumptions 4.1, 4.3 and 4.5 hold. Then $\hat{X}_{\mathcal{T}}$ has a single recurrent class. Moreover, if Assumption 4.4 holds, the process $\hat{X}_{\mathcal{T}}$ is irreducible.*

Proof. Let $\mathcal{D} \subseteq \mathcal{T}$ be a non-empty recurrent class for $\hat{X}_{\mathcal{T}}$ (there must be at least one), and let $\mathcal{C} \subseteq \mathcal{X}$ be the communicating class under $\tilde{X}$ described in Assumption 4.3. We will prove that $\mathcal{D} = \mathcal{C} \setminus \mathcal{A}$, which yields the uniqueness of recurrent classes for $\hat{X}_{\mathcal{T}}$. If Assumption 4.4 holds, then $\mathcal{C} = \mathcal{X}$, which combined with the relation $\mathcal{D} = \mathcal{C} \setminus \mathcal{A}$, implies that $\mathcal{D} = \mathcal{X} \setminus \mathcal{A} = \mathcal{T}$ and the conclusion follows.

In order to prove $\mathcal{D} = \mathcal{C} \setminus \mathcal{A}$, we start by making some observations. First, we prove that there exist states $\tilde{x} \in \mathcal{D}$ and $\tilde{y} \in \mathcal{A}$ such that $\tilde{Q}_{\tilde{x},\tilde{y}} > 0$. In fact, if this was not the case, then for every $x \in \mathcal{D}$ and $z \in \mathcal{A}$ we would have $\tilde{Q}_{x,z} = (R_0)_{x,z} = 0$. This yields that for $x \in \mathcal{D}$,

$$(Q_{\mathcal{T}})_{x,y} = (T_0)_{x,y} + \sum_{z\in\mathcal{A}} (R_0)_{x,z}[(-A_1)^{-1}S_1]_{z,y} = (T_0)_{x,y}, \tag{S.19}$$

for all $y \in \mathcal{T}$. Since $\mathcal{D}$ is a closed class under $Q_{\mathcal{T}}$, $(Q_{\mathcal{T}})_{x,y} = 0$ for $y \in \mathcal{T} \setminus \mathcal{D}$ and so $\sum_{y\in\mathcal{D}}(Q_{\mathcal{T}})_{x,y} = \sum_{y\in\mathcal{T}}(Q_{\mathcal{T}})_{x,y} = 0$, since $Q_{\mathcal{T}}$ is a Q-matrix. Combining this with the previous equation, we obtain that $\sum_{y\in\mathcal{D}}(T_0)_{x,y} = 0$ for all $x \in \mathcal{D}$, which implies that $\mathcal{D}$ is closed under $\tilde{Q}$. This contradicts the fact that T_0 is invertible by point (i) in Lemma S.5 (with $\mathcal{B}^c = \mathcal{T}$ and $Q = \tilde{Q}$).

Second, we observe that there exist a $\hat{y} \in \mathcal{A}$ and $\hat{x} \in \mathcal{D}$ such that $\tilde{Q}_{\hat{y},\hat{x}} > 0$. In fact, since A_1 is invertible, by Lemma S.5 (with $\mathcal{B}^c = \mathcal{A}$ and $Q = \tilde{Q}$) there has to be a $\hat{y} \in \mathcal{A}$ and $\hat{x} \in \mathcal{T}$ such that $\tilde{Q}_{\hat{y},\hat{x}} > 0$. To show that $\hat{x} \in \mathcal{D}$, consider that $\tilde{y} \in \mathcal{A}$ communicates with $\hat{y} \in \mathcal{A}$ under $\tilde{Q}$, by Assumption 4.3, and therefore $\tilde{x}$ leads to $\hat{x}$ under $\tilde{Q}$ and therefore under $Q_{\mathcal{T}}$. Since $\mathcal{D}$ is closed under $Q_{\mathcal{T}}$, $\hat{x} \in \mathcal{D}$.

We now prove that $\mathcal{D} \subseteq \mathcal{C} \setminus \mathcal{A}$. For $x \in \mathcal{D}$, since T_0 is invertible there exists a state $y \in \mathcal{A}$ such that x leads to y under $\tilde{Q}$. By Assumption 4.3, y and $\hat{y}$ are in $\mathcal{A} \subseteq \mathcal{C}$ and so they communicate under $\tilde{Q}$. It follows that x leads to $\hat{y}$ under $\tilde{Q}$. On the other hand, $\tilde{Q}_{\hat{y},\hat{x}} > 0$ and since $\mathcal{D}$ is a communicating class under $Q_{\mathcal{T}}$, $\hat{x}$ leads to x under $\tilde{Q}$. Thus, x leads to $\hat{y}$ and $\hat{y}$ leads to x under $\tilde{Q}$ and so $x \in \mathcal{C}$. Thus, $\mathcal{D} \subseteq \mathcal{C}$ and $\mathcal{D} \subseteq \mathcal{T} = \mathcal{A}^c$, and so $\mathcal{D} \subseteq \mathcal{C} \setminus \mathcal{A}$.

To prove that $\mathcal{C} \setminus \mathcal{A} \subseteq \mathcal{D}$, let $x \in \mathcal{C} \setminus \mathcal{A}$. Since $\mathcal{D} \subseteq \mathcal{C}$, then x communicates with the element $\tilde{x}$ in $\mathcal{D}$ under $\tilde{Q}$. This implies that they communicate under $Q_{\mathcal{T}}$ and since $\mathcal{D}$ is a communicating class under $Q_{\mathcal{T}}$, then $x \in \mathcal{D}$. Combining the above we see that $\mathcal{D} = \mathcal{C} \setminus \mathcal{A}$. ∎

When the continuous time Markov chain $\hat{X}_{\mathcal{T}}$ has a single recurrent class $\mathcal{D}$, there is a unique probability vector ν in $\mathbb{R}^{|\mathcal{T}|}$ such that $\nu Q_{\mathcal{T}} = 0$ and ν will be the stationary distribution for $\hat{X}_{\mathcal{T}}$ with non-zero entries only for entries corresponding to states in $\mathcal{D}$. We use the vector ν to characterize α and $\beta^{(1)}$.

In the following theorem, we use the fact that A_1 is invertible. This follows from Lemma S.4 because $A_1 = Q^{\mathcal{A}}(1)$, where $Q(1)$ is positive recurrent and so the condition (i) of Lemma S.4 holds with $B_c = \mathcal{A}$.

Theorem S.11. *Suppose Assumptions 4.1, 4.3 and 4.5 hold. Denote by ν the unique probability vector in $\mathbb{R}^{|\mathcal{T}|}$ such that $\nu Q_{\mathcal{T}} = 0$. Then, $\pi(0) = [\alpha, 0]$ where α is given by*

$$\alpha = c\nu R_0(-A_1)^{-1}, \tag{S.20}$$

and where c is given by $c = (\nu R_0(-A_1)^{-1}\mathbb{1})^{-1}$. Moreover, $\pi^{(1)} = [\alpha^{(1)}, \beta^{(1)}]$ where

$$\beta^{(1)} = c\nu. \tag{S.21}$$

Proof. Following the proof of Theorem S.9, equations (S.14) and (S.15) yield that

$$\beta^{(1)} R_0 + \alpha A_1 = 0, \tag{S.22}$$

and

$$\beta^{(1)} T_0 + \alpha S_1 = 0. \tag{S.23}$$

From (S.22) we obtain that $\alpha = \beta^{(1)} R_0(-A_1)^{-1}$. We substitute this expression in (S.23) to obtain that $\beta^{(1)}(T_0 + R_0(-A_1)^{-1}S_1) = 0$, which is exactly $\beta^{(1)} Q_{\mathcal{T}} = 0$. By Lemma S.10 combined with Lemma S.1, we obtain that $\beta^{(1)} = \tilde{c}\nu$ for some constant $\tilde{c} \in \mathbb{R}$ and therefore $\alpha = \tilde{c}\nu R_0(-A_1)^{-1}$. To show that $\tilde{c} = c$, we observe that since $\alpha\mathbb{1} = 1$, then $\tilde{c}(\nu R_0(-A_1)^{-1}\mathbb{1}) = 1$ and the desired result follows. ∎

Under the assumptions of Theorem S.11, $\beta_x^{(1)} > 0$ for every $x \in \mathcal{D}$, the single recurrent class of $\hat{X}_{\mathcal{T}}$, while $\beta_x^{(1)} = 0$ for $x \in \mathcal{T}^{\backslash(1)}$. In fact, using first step analysis, one can show that the entry $(-A_1)^{-1}_{i,j}$ is the expected time that the process $\tilde{X}$ spends at j when started at i, before exiting $\mathcal{A}$. Hence, these entries are non-negative and so does $\nu R_0(-A_1)^{-1}$. This implies that the constant c is positive and the conclusion follows from (S.21) and the properties of ν.

S.2.4 Additional characterization of zeroth and first order terms via partial balance

For the last part of this section, we consider an additional characterization for $\beta^{(1)}$ based on the idea of truncated processes and partial balance relations (see Section 9.4 in [13]).

Consider a continuous time Markov chain $X = \{X(t) : t \geq 0\}$ with infinitesimal generator Q on a finite state space $\mathcal{X}$. Let $\mathcal{Y}$ a non-empty set in $\mathcal{X}$. Define the matrix $\bar{Q} = (\bar{Q}_{x,y})_{x,y\in\mathcal{Y}}$ by $\bar{Q}_{x,y} = Q_{x,y}$ for $x \neq y$ and $\bar{Q}_{x,x} = Q_{x,x} + \sum_{y\notin\mathcal{Y}} Q_{x,y}$. A continuous time Markov chain $\bar{X} = \{\bar{X}(t) : t \geq 0\}$ with state space $\mathcal{Y}$ and infinitesimal generator $\bar{Q}$ will be called a **truncation** of X to $\mathcal{Y}$.

Assumption S.1. *For every $0 < \varepsilon < \varepsilon_0$, the truncation of X^ε to $\mathcal{T}$, denoted by $\bar{X}^\varepsilon$, is irreducible. In addition, the following partial balance condition holds on $\mathcal{T}$ for every $0 < \varepsilon < \varepsilon_0$:*

$$\pi_x(\varepsilon) \sum_{y\in\mathcal{A}} Q_{x,y}(\varepsilon) = \sum_{y\in\mathcal{A}} \pi_y(\varepsilon) Q_{y,x}(\varepsilon), \qquad \text{for every } x \in \mathcal{T}. \tag{S.24}$$

Under Assumption S.1, the process $\bar{X}^\varepsilon$ has a stationary distribution $\eta(\varepsilon)$ for every $0 < \varepsilon < \varepsilon_0$, given by

$$\eta_x(\varepsilon) = \frac{\pi_x(\varepsilon)}{\sum_{y\in\mathcal{T}} \pi_y(\varepsilon)}, \quad x \in \mathcal{T} \tag{S.25}$$

(see Theorem 9.5 in Kelly [13]). The following is our main theorem.

Theorem S.12. *Suppose Assumptions 4.1, 4.5 and S.1 hold. Then, the following hold:*

(i) the limit $\eta := \lim_{\varepsilon\to 0} \eta(\varepsilon)$ exists and it is a probability vector on $\mathcal{T}$ such that

$$\eta \bar{Q}(0) = 0, \tag{S.26}$$

(ii) the vectors α and $\beta^{(1)}$ can be characterized by

$$\beta^{(1)} = c\eta, \qquad \alpha = c\eta R_0(-A_1)^{-1}, \tag{S.27}$$

where $c = (\eta R_0(-A_1)^{-1}\mathbb{1})^{-1}$, and

(iii) $\eta Q_{\mathcal{T}} = 0$.

If, in addition, Assumption 4.3 or Assumption 4.4 holds, then $\eta = \nu$ is the unique stationary distribution for $\hat{X}_{\mathcal{T}}$.

Remark S.13. Although we know that $\bar{X}^0$ is well defined, we do not know a priori whether the process is irreducible or it has a single recurrent class. If the truncation process $\bar{X}^0$ has a single recurrent class (or is irreducible), it will have a unique stationary distribution, which we would call $\eta(0)$. But we do not know if such a vector exists. This non existence is what led us to express Theorem S.12 in terms of the limit η which solves $\eta\bar{Q}(0) = 0$. If the truncation process $\bar{X}^0$ has a single recurrent class, as in the 1D and 2D models, the probability vector η is characterized uniquely by solving $\eta\bar{Q}(0) = 0$ and $\eta = \eta(0)$.

Proof. We will first show that $\beta^{(1)}\mathbb{1} > 0$. From (4.5) we know that $\beta^{(1)} = \alpha S_1(-T_0)^{-1}$, which yields $\beta^{(1)}\mathbb{1} = \alpha S_1(-T_0)^{-1}\mathbb{1}$. Since all of the entries in α, S_1, and $(-T_0)^{-1}$ are nonnegative, it suffices to show $\beta^{(1)}\mathbb{1} \neq 0$. For a proof by contradiction, suppose that $\beta^{(1)}\mathbb{1} = 0$. This means that

$$\sum_{y\in\mathcal{T}} \sum_{x\in\mathcal{A}} \alpha_x (S_1)_{x,y} ((-T_0)^{-1}\mathbb{1})_y = 0. \tag{S.28}$$

All of the entries in the sum are nonnegative, so this means that $\alpha_x(S_1)_{x,y}((-T_0)^{-1}\mathbb{1})_y = 0$ for every $x \in \mathcal{A}$ and $y \in \mathcal{T}$. Now, $((-T_0)^{-1}\mathbb{1})_y$ is the mean first passage time to $\mathcal{A}$, for the Markov chain that starts at y with infinitesimal generator $Q(0)$, and so $((-T_0)^{-1}\mathbb{1})_y \geq \frac{1}{|Q(0)_{y,y}|} > 0$.

Hence, $\alpha_x(S_1)_{x,y} = 0$ for every $x \in \mathcal{A}$ and $y \in \mathcal{T}$. This yields that $\alpha S_1 = 0$ and substituting this in (4.4) yields that $\alpha A_1 = 0$. Since A_1 is invertible, this is a contradiction.

Since we know that $\beta(0) = \beta^{(0)} = 0$, we obtain that

$$\eta_x(\varepsilon) = \frac{\pi_x(\varepsilon)}{\sum_{y\in\mathcal{T}} \pi_y(\varepsilon)} = \frac{\sum_{k=1}^{\infty} \varepsilon^k \beta_x^{(k)}}{\sum_{y\in\mathcal{T}} \sum_{k=1}^{\infty} \varepsilon^k \beta_y^{(k)}} = \frac{\sum_{k=1}^{\infty} \varepsilon^k \beta_x^{(k)}}{\sum_{k=1}^{\infty} \varepsilon^k \sum_{y\in\mathcal{T}} \beta_y^{(k)}} = \frac{\sum_{k=1}^{\infty} \varepsilon^{k-1} \beta_x^{(k)}}{\sum_{k=1}^{\infty} \varepsilon^{k-1} \sum_{y\in\mathcal{T}} \beta_y^{(k)}}$$
$$\to \frac{\beta_x^{(1)}}{\sum_{y\in\mathcal{T}} \beta_y^{(1)}} = \frac{\beta_x^{(1)}}{\beta^{(1)}\mathbb{1}}.$$

We then obtain that η exists and $\eta_x = \frac{\beta_x^{(1)}}{\beta^{(1)}\mathbb{1}}$ for every $x \in \mathcal{T}$, which is a probability vector on $\mathcal{T}$. Or letting $\varepsilon \to 0$ in $\eta(\varepsilon)\bar{Q}(\varepsilon) = 0$, we obtain that $\eta\bar{Q} = 0$. We already know that $\beta^{(1)} = c\eta$. To obtain a value for c that depends only on η, note that from (4.4) and (4.5), we have $\alpha = \beta^{(1)}R_0(-A_1)^{-1} = c\eta R_0(-A_1)^{-1}$, where $c\eta R_0(-A_1)^{-1}\mathbb{1} = 1$ and so $c = (\eta R_0(-A_1)^{-1}\mathbb{1})^{-1}$.

By following the proof of Theorem S.11, we obtain $\beta^{(1)}Q_{\mathcal{T}} = 0$ and therefore $\eta Q_{\mathcal{T}} = 0$. The other conclusions follow readily. ■

The following criterion offers a practical way to establish (S.24).

Lemma S.14. *Let $\mathcal{A} = \{a_1, \ldots, a_n\}$. Suppose there exist distinct states $x_1, \ldots, x_n$ in $\mathcal{T}$ such that for every $0 < \varepsilon < \varepsilon_0$ and for every $k \in \{1, \ldots, n\}$.*

1. *$Q_{a_k x_k}(\varepsilon), Q_{x_k a_k}(\varepsilon) > 0$,*
2. *$Q_{a_k y}(\varepsilon) = Q_{y a_k}(\varepsilon) = 0$ for every $y \notin \{x_k, a_k\}$.*

Then (S.24) holds.

Proof. Denote by $\mathcal{N} = \{x_1, \ldots, x_n\}$. Let $0 < \varepsilon < \varepsilon_0$. For $x \in \mathcal{T} \setminus \mathcal{N}$, we have that $\pi_x(\varepsilon)\sum_{y\in\mathcal{A}} Q_{x,y}(\varepsilon) = 0$ and $\sum_{y\in\mathcal{A}} \pi_y(\varepsilon)Q_{y,x}(\varepsilon) = 0$. Then, equation (S.24) holds for $x \in \mathcal{T} \setminus \mathcal{N}$.

For $x_k \in \mathcal{N}$, $\pi_{x_k}(\varepsilon)\sum_{y\in\mathcal{A}} Q_{x_k,y}(\varepsilon) = \pi_{x_k}(\varepsilon)Q_{x_k a_k}(\varepsilon)$.

On the other hand, $\sum_{y\in\mathcal{A}} \pi_y(\varepsilon)Q_{y x_k}(\varepsilon) = \pi_{a_k}(\varepsilon)Q_{a_k x_k}(\varepsilon)$. Since $\pi(\varepsilon)Q(\varepsilon) = 0$, we have

$$0 = (\pi(\varepsilon)Q(\varepsilon))_{a_k} = \sum_{x\in\mathcal{X}} \pi_x(\varepsilon)Q_{x,a_k}(\varepsilon)$$
$$= \pi_{x_k}(\varepsilon)Q_{x_k,a_k}(\varepsilon) + \pi_{a_k}(\varepsilon)Q_{a_k,a_k}(\varepsilon) = \pi_{x_k}(\varepsilon)Q_{x_k,a_k}(\varepsilon) - \pi_{a_k}(\varepsilon)Q_{a_k,x_k}(\varepsilon).$$

Hence, $\pi_{x_k}(\varepsilon)Q_{x_k,a_k}(\varepsilon) = \pi_{a_k}(\varepsilon)Q_{a_k,x_k}(\varepsilon)$ and (S.24) holds for $x \in \mathcal{N}$ as well. ■

S.2.5 Lemma S.15

Lemma S.15. *Under Assumption 4.1, the matrices $Q_{\mathcal{A}}$ and $\tilde{Q}$ are Q-matrices of sizes $|\mathcal{A}|\times|\mathcal{A}|$ and $|\mathcal{X}|\times|\mathcal{X}|$ respectively. If in addition, Assumption 4.5 holds, then $Q_{\mathcal{T}}$ is a Q-matrix of size $|\mathcal{T}|\times|\mathcal{T}|$.*

Proof. First, observe that $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}Q_{x,y}(\varepsilon) = (A_1)_{x,y}$ if $x, y \in \mathcal{A}$, while $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}Q_{x,y}(\varepsilon) = (S_1)_{x,y}$ if $x \in \mathcal{A}$ and $y \in \mathcal{T}$. Then, since $Q(\varepsilon)$ is a Q-matrix, S_1 has nonnegative entries, $(A_1)_{x,y} \geq 0$ for $x \neq y \in \mathcal{A}$, and

$$\sum_{y\in\mathcal{A}} (A_1)_{x,y} + \sum_{y\in\mathcal{T}} (S_1)_{x,y} = 0 \text{ for every } x \in \mathcal{A}. \quad \text{(S.29)}$$

For $x \neq y \in \mathcal{A}$, $(Q_{\mathcal{A}})_{x,y} = (A_1)_{x,y} + \sum_{z\in\mathcal{T}}(S_1)_{x,z}((-T_0)^{-1}R_0)_{z,y}$ is nonnegative since $(A_1)_{x,y} \geq 0$ and by (S.2), $((-T_0)^{-1}R_0)_{z,y} = 0$ for $z \in \mathcal{T}$. For $x \in \mathcal{A}$, $\sum_{y\in\mathcal{A}}(Q_{\mathcal{A}})_{x,y}$ is equal to

$$\sum_{y\in\mathcal{A}}(A_1)_{x,y} + \sum_{y\in\mathcal{A}}\sum_{z\in\mathcal{T}}(S_1)_{x,z}((-T_0)^{-1}R_0)_{z,y} = \sum_{y\in\mathcal{A}}(A_1)_{x,y} + \sum_{z\in\mathcal{T}}(S_1)_{x,z}\sum_{y\in\mathcal{A}}((-T_0)^{-1}R_0)_{z,y}$$
$$= \sum_{y\in\mathcal{A}}(A_1)_{x,y} + \sum_{z\in\mathcal{T}}(S_1)_{x,z} = 0,$$

where we used (S.2) and (S.29). Hence $Q_{\mathcal{A}}$ is a Q-matrix.

For $\tilde{Q}$, for $x \neq y \in \mathcal{X}$, if $x \in \mathcal{A}$, then $\tilde{Q}_{x,y}$ corresponds to an off diagonal term in A_1 or a term in S_1, both of which are nonnegative. If $x \in \mathcal{T}$, then $\tilde{Q}_{x,y}$ corresponds to an off diagonal term in T_0 or a term in R_0, which are both nonegative since $Q(0)$ is a Q-matrix. To check that the row-sums of $\tilde{Q}$ are zero, first consider when $x \in \mathcal{A}$. Then, $\sum_{y\in\mathcal{X}}\tilde{Q}_{x,y} = \sum_{y\in\mathcal{A}}(A_1)_{x,y} + \sum_{y\in\mathcal{T}}(S_1)_{x,y} = 0$ by (S.29). If $x \in \mathcal{T}$, then $\sum_{y\in\mathcal{X}}\tilde{Q}_{x,y} = \sum_{y\in\mathcal{A}}(R_0)_{x,y} + \sum_{y\in\mathcal{T}}(T_0)_{x,y} = 0$, since this corresponds to summing across a row of $Q(0)$.

The case of $Q_{\mathcal{T}}$ follows similarly to that for $Q_{\mathcal{A}}$. ■

S.3 Algorithm to find the order of the pole of the MFPT

Input: $\mathcal{B} \subset \mathcal{X}$, and k_{xy}, the order of $Q_{x,y}(\varepsilon)$ for each $(x, y) \in E_0$.
Output: $p(x)$, the order of the pole of the mean first passage from $x \in \mathcal{B}^c$ to $\mathcal{B}$.
(p will also be defined for condensed nodes in the course of the algorithm)
Step 1 (Set up the initial graph (V, E))
Construct a directed graph $G = (V, E)$ where $V = \mathcal{X}$ and $E = E_0$.
Set, for each $u \in V$, $p(u) \leftarrow \min\{k_{uv} : (u, v) \in E\}$.
Set, for each $(u, v) \in E$, $\mathcal{K}_{uv} \leftarrow k_{uv} - p(u)$.
Step 2 (Condense $\mathcal{B}$ into a single node a)
Introduce a new node a.
Set, for each $w \in \mathcal{B}^c$ such that $(w, v) \in E$ for some $v \in \mathcal{B}$, $\mathcal{K}_{wa} \leftarrow \min\{\mathcal{K}_{wv} : v \in \mathcal{B} \text{ and } (w, v) \in E\}$.
Update $V \leftarrow \mathcal{B}^c \cup \{a\}$ and
$E \leftarrow \{(u, v) \in E : u \in \mathcal{B}^c \text{ and } v \in \mathcal{B}^c\} \cup \{(w, a) : (w, v) \in E \text{ for some } w \in \mathcal{B}^c \text{ and } v \in \mathcal{B}\}$.
Set, for each $u \in V \setminus \{a\}$, $S(u) \leftarrow \{u\}$, and $S(a) \leftarrow \mathcal{B}$.
Step 3 (Condense r-connected sets)
Repeat the following until G contains no r-connected sets:
Let $C \subset V$ be an r-connected set and c be a new node representing the r-connected set C.
Set $p(c) \leftarrow \max_{u\in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\}$.
Set, for each $w \in V \setminus C$ such that $(u, w) \in E$ for some $u \in C$,
$\mathcal{K}_{cw} \leftarrow \min\{\mathcal{K}_{uw} : u \in C \text{ and } (u, w) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\}$
Set, for each $w \in V \setminus C$ such that $(w, v) \in E$ for some $v \in C$,
$\mathcal{K}_{wc} \leftarrow \min\{\mathcal{K}_{wv} : v \in C \text{ and } (w, v) \in E\}$.
Update $V \leftarrow (V \setminus C) \cup \{c\}$ and
$E \leftarrow \{(u, v) \in E : u \notin C \text{ and } v \notin C\} \cup \{(c, w) : (u, w) \in E \text{ for some } u \in C \text{ and } w \notin C\}$
$\cup\{(w, c) : (w, v) \in E \text{ for some } w \notin C \text{ and } v \in C\}$.
Set $S(c) \leftarrow \cup_{u\in C} S(u)$.
Step 4 (Compute $p(x)$ where $x \in \mathcal{B}^c$)
Repeat the following until $V = \{a\}$:
Let $v^* \in V \setminus \{a\}$ be such that $p(v^*) = \max_{u\in V\setminus\{a\}} p(u)$ and break the tie arbitrarily.

For each $x \in S(v^*) \subset \mathcal{B}^c$, the order of the pole of the mean first passage time from x to $\mathcal{B}$ is $p(x) \leftarrow p(v^*)$.

Update, for each $u \in V$ such that $(u, v^*) \in E$, $p(u) \leftarrow \max\{p(u), p(v^*) - \mathcal{K}_{uv^*}\}$.

Update $V \leftarrow V \setminus \{v^*\}$ and $E \leftarrow \{(u, v) \in E : u \neq v^* \text{ and } v \neq v^*\}$.

S.4 Graphs for the algorithm to find the order of the MFPT

Here we elaborate on the graphs that we use in the algorithm and the definitions that we gave in Section 4.2.1 and Section S.3. While, in the algorithm statement, we used the same notation for the updated graphs as in the original graph, it will be clearer for the justification given in Section S.5 if we specify which copy of the graph we are looking at for each step. Accordingly, we provide a more detailed version of the definitions of these graphs and associated notation in this section.

Step 1: Graph G

For each $\varepsilon \in (0, \varepsilon_{\max})$ and $x \in \mathcal{X}$, the exponential parameter satisfies

$$q_x(\varepsilon) = -Q_{x,x}(\varepsilon) = \sum_{y \neq x \in X} Q_{x,y}(\varepsilon) = \sum_{(x,y) \in E_0} Q_{x,y}(\varepsilon) > 0.$$

Since the order of $Q_{x,y}(\varepsilon)$ is k_{xy} for each $y \in \mathcal{X}$ such that $(x, y) \in E_0$, the order of $q_x(\varepsilon)$ is $p_0(x) = \min\{k_{xy} : (x, y) \in E_0\}$. For each $(x, y) \in E_0$, the transition probability $P_{x,y}(\varepsilon)$ for the embedded discrete time Markov chain is

$$P_{x,y}(\varepsilon) = \frac{Q_{x,y}(\varepsilon)}{q_x(\varepsilon)},$$

the order of which is

$$\mathcal{K}_{xy} = k_{xy} - \min\{k_{xy} : (x, y) \in E_0\} = k_{xy} - p_0(x). \tag{S.30}$$

We start with a weighed graph $G = (V, E)$ where $V = \mathcal{X}$ and $E = E_0$. For each $u \in V$, the node weight of u is $p_0(u)$, which is the **order of the pole of the expected sojourn time** at state x until escape from x for X^ε. For each $(u, v) \in E$, the edge weight of (u, v) is $\mathcal{K}_{uv}$, which is the **order of the transition probability** from u to v.

Step 2: Graph $G^{(0)}$

If $x \in \mathcal{B}^c$ is such that $(x, y) \in E$ for some $y \in \mathcal{B}$, then the transition probability from x to $\mathcal{B}$ is positive and is given by

$$P_{x,\mathcal{B}}(\varepsilon) = \sum_{\substack{y \in \mathcal{B}: \\ (x,y) \in E}} P_{x,y}(\varepsilon).$$

For such x, the order of $P_{x,\mathcal{B}}(\varepsilon)$ is

$$\mathcal{K}_{xa} = \min\{\mathcal{K}_{xy} : y \in \mathcal{B}, (x, y) \in E\}, \tag{S.31}$$

where $\mathcal{K}_{xy}$ is the order of $P_{x,y}(\varepsilon)$ for each $(x, y) \in E$.

Now, we are ready to specify the graph $G^{(0)} = (V^{(0)}, E^{(0)})$ which serves as the base case for Step 3. We group the nodes in $\mathcal{B}$ into a single node, denoted by a, so the set of nodes becomes $V^{(0)} = (V \setminus \mathcal{B}) \cup \{a\} = \mathcal{B}^c \cup \{a\}$. All of the edges starting from or going to a node in $\mathcal{B}$ are then removed. If there was an edge from $x \in \mathcal{B}^c$ to a node in $\mathcal{B}$, then we add back an edge (x, a). We leave out all edges from a node in $\mathcal{B}$ to a node in $\mathcal{B}^c$, since we are interested in the mean first passage time to the set $\mathcal{B}$. The resulting edge set is $E^{(0)} = \{(u, v) \in E : u \in \mathcal{B}^c \text{ and } v \in \mathcal{B}^c\} \cup \{(w, a) : (w, v) \in E \text{ for some } w \in \mathcal{B}^c \text{ and } v \in \mathcal{B}\}$.

Let $S(u) = \{u\}$ for all $u \in V^{(0)} \setminus \{a\}$ and $S(a) = \mathcal{B}$. Note that $\{S(u) : u \in V^{(0)}\}$ is a partition of the state space $\mathcal{X}$, denoting the grouping of nodes in $V^{(0)}$. For each $u \in V^{(0)} \setminus \{a\}$ and $x \in S(u) = \{u\}$, we define $p_0^x(u)$ to be the **order of the pole of the expected sojourn time** in $S(u)$ before exiting $S(u)$ when starting the process at the state x. For each $(u, v) \in E^{(0)}$ and $x \in S(u) = \{u\}$, we define $\mathcal{K}_{uv}^x$ to be the **order of the probability of a transition** to $S(v)$ upon exiting from $S(u)$ when the process is started at the state x. For these terms, $p_0^x(u) = p_0(u)$ and $\mathcal{K}_{uv}^x = \mathcal{K}_{uv}$, which is the base case for Lemma S.20.

Step 3: Graphs $\{G^{(N)}\}_{N=0}^M$

In Step 3, we define a sequence of graphs $\{G^{(N)} = (V^{(N)}, E^{(N)})\}_{N=0}^M$ recursively, where the exact value of $M \geq 0$ is not pre-determined and is only revealed when an exit condition for the recursion is satisfied. We know this recursion will end after a finite number of iterations because the number of nodes in $V^{(N)}$ is strictly decreasing with N. The weight p_0 of each node and the weight $\mathcal{K}$ of each edge are also defined iteratively, and each is defined only once.

We have already defined $G^{(0)}$ in Step 2. Fix $N \in \{1, 2, \ldots, M+1\}$, where the value of $M < \infty$ is defined below. At the N^{th} iteration, an edge $(u, v) \in E^{(N-1)}$ is called an **r-edge** if its edge weight $\mathcal{K}_{uv}$ is 0; a directed path in $G^{(N-1)}$ is called an **r-path** if it consists of r-edges only. A set $C \subset V^{(N-1)}$ is called an **r-connected set** in $G^{(N-1)}$ if $|C| > 1$ and there exists an r-path from u to v for any $u \neq v \in C$. Here we use the qualifier "r" to indicate that these edges, paths and cycles are "regular". If there is no r-connected set in $G^{(N-1)}$, the iteration stops. We set the value of M to the first value of $N-1$ such that $G^{(N-1)}$ does not have any r-connected set. At that time point, the iteration stops and we move to Step 4 where $G^{(M)}$ will be the initial graph for Step 4. Otherwise, $N \in \{1, \ldots, M\}$, and we let C_N be an r-connected set in $G^{(N-1)}$, which is condensed to a new node c_N in $G^{(N)}$. Then, we define the graph $G^{(N)} = (V^{(N)}, E^{(N)})$, where $V^{(N)} = (V^{(N-1)} \setminus C_N) \cup \{c_N\}$, and $E^{(N)} = \{(u, v) \in E^{(N-1)} : u \notin C_N \text{ and } v \notin C_N\} \cup \{(c_N, w) : (u, w) \in E^{(N-1)} \text{ for some } u \in C_N \text{ and } w \notin C_N\} \cup \{(w, c_N) : (w, v) \in E^{(N-1)} \text{ for some } w \notin C_N \text{ and } v \in C_N\}$. Let $S(c_N) = \cup_{u \in C_N} S(u)$. Note that $\{S(u) : u \in V^{(N)}\}$ is again a partition of the state space $\mathcal{X}$, denoting the grouping of nodes in $V^{(N)}$.

We define $p_0^x(c_N)$ to be the **order of the pole of the expected sojourn time** in $S(c_N)$ until the first exit from $S(c_N)$ when the process is started at the state $x \in S(c_N)$. In Lemma S.20, we will show that the value of $p_0^x(c_N)$ is independent of the state $x \in S(c_N)$, and we define $p_0(c_N) = p_0^x(c_N)$. For each $w \in V^{(N-1)} \setminus C_N$ such that $(w, c_N) \in E^{(N)}$, define $\mathcal{K}_{wc_N}^x$ to be the **order of the probability of a transition** to $S(c_N)$ upon exiting from $S(w)$ when the process is started at the state $x \in S(w)$. In Lemma S.20, we will show that the value of $\mathcal{K}_{wc_N}^x$ is independent of the state $x \in S(w)$, and $\mathcal{K}_{wc_N}^x = \mathcal{K}_{wc_N}$ where

$$\mathcal{K}_{wc_N} = \min\{\mathcal{K}_{wv} : v \in C_N \text{ and } (w, v) \in E^{(N-1)}\}. \tag{S.32}$$

For each $w \in V^{(N-1)} \setminus C_N$ such that $(c_N, w) \in E^{(N)}$, define $\mathcal{K}_{c_N w}^x$ to be the **order of the probability of a transition** to $S(w)$ upon exiting from $S(c_N)$ when the process is started at the state $x \in S(c_N)$. In Lemma S.20, we will show that the value of $\mathcal{K}^x_{c_N w}$ is independent of the state $x \in S(c_N)$, and $\mathcal{K}^x_{c_N w} = \mathcal{K}_{c_N w}$ where

$$\begin{aligned} \mathcal{K}_{c_N w} &= \min\{\mathcal{K}_{uw} : u \in C_N \text{ and } (u,w) \in E^{(N-1)}\} \\ &\quad - \min\{\mathcal{K}_{uv} : u \in C_N, v \notin C_N \text{ and } (u,v) \in E^{(N-1)}\}. \end{aligned} \tag{S.33}$$

We note that for each $N \in \{1,\ldots,M\}$, since there is no edge in $E^{(N-1)}$ that leads from a, the node a is never part of any r-connected set, and so $a \in V^{(N)}$ and there is at least one other node in $V^{(N)}$ besides a. Also, the irreducibility of X^ε when $0 < \varepsilon < \varepsilon_0$ implies that there is a path from x to y in G for each $x \in \mathcal{X} \setminus \mathcal{B}$ and $y \in \mathcal{X}$. This implies that if $u_N \neq v_N \in V^{(N)}$ for some $N \in \{0,1,\ldots,M\}$ such that $x \in S(u_N)$ and $y \in S(v_N)$, then there is a path from u_N to v_N in $G^{(N)}$. Therefore, for each $N \in \{0,1,\ldots,M\}$, there is always an outgoing edge from some $u' \in C_N$ (u' cannot be a) to some $v' \in V^{(N-1)} \setminus C_N$ in $G^{(N-1)}$. In addition, as can be seen from the definition of the $\mathcal{K}$'s in (S.30), (S.31), (S.32) and (S.33), for each $N = 0,1,\ldots,M$ and $u'' \in V^{(N)} \setminus \{a\}$, there exists an r-edge $(u'',v'') \in E^{(N)}$ for some $v'' \in V^{(N)}$. For $G^{(M)}$, $|V^{(M)}| \geq 2$. Furthermorer, if we only look at r-edges (and ignore the other edges), $G^{(M)}$ is an acyclic graph (as it contains no r-connected set). It follows that the node a is the only sink because for each $u \in V^{(M)} \setminus \{a\}$, there is an outgoing r-edge, and thus there is an r-path from u to a for each $u \in V^{(M)} \setminus \{a\}$.

Step 4: Graphs $\{G^{(M,N)}\}_{N=0}^{|V^{(M)}|-1}$

In Step 4, we define a sequence of graphs $\{G^{(M,N)} = (V^{(M,N)}, E^{(M,N)})\}_{N=0}^{|V^{(M)}|-1}$ recursively, where $G^{(M,0)} = G^{(M)}$. In each iteration, the weight of one of the nodes in $V^{(M,N-1)} \subset V^{(M)}$ is finalized and determines the value of p there, and the weights p_{N-1} of other nodes in $V^{(M,N)}$ are updated to p_N.

Fix $N \in \{1,\ldots,|V^{(M)}|-1\}$. At the N^{th} iteration, let $v_N \in V^{(M,N-1)} \setminus \{a\}$ be such that

$$p_{N-1}(v_N) = \max_{u \in V^{(M,N-1)} \setminus \{a\}} p_{N-1}(u) =: p(v_N), \tag{S.34}$$

where we break the tie arbitrarily.

Now, we define the graph $G^{(M,N)} = (V^{(M,N)}, E^{(M,N)})$ where $V^{(M,N)} = V^{(M,N-1)} \setminus \{v_N\}$ and $E^{(M,N)} = \{(u,v) \in E^{(M,N-1)} : u \neq v_N \text{ and } v \neq v_N\}$. For each $u \in V^{(M,N)}$, let

$$p_N(u) = \begin{cases} \max\{p_{N-1}(u), p_{N-1}(v_N) - \mathcal{K}_{uv_N}\}, & \text{for } (u,v_N) \in E^{(M,N-1)}, \\ p_{N-1}(u), & \text{for } (u,v_N) \notin E^{(M,N-1)}. \end{cases} \tag{S.35}$$

In Theorem S.22, we will show that for each $x \in S(v_N)$, the order of the pole of the mean first passage time from x to the set $\mathcal{B}$ is $p(x) = p(v_N)$.

S.5 Justification for the algorithm to find the order of the MFPT Recall that we defined $\{G^{(N)}\}_{N=0}^M = \{(V^{(N)}, E^{(N)})\}_{N=0}^M$ and $G^{(M,0)} = G^{(M)}$ in Section S.4. Each $G^{(N)}$ defines a partition $\{S(u) : u \in V^{(N)}\}$ of $\mathcal{X}$, which will be used in our proofs. Since X^ε is an irreducible continuous time Markov chain for all $\varepsilon \in (0,\varepsilon_0)$, each $G^{(N)}$ is weakly connected and has the property that any node in $V^{(N)} \setminus \{a\}$ has an out-going edge starting from the node.

In this section, we will provide the justification for the algorithm. Step 1 of the algorithm sets up the original continuous time Markov chain using a skeleton chain. Step 2 of the algorithm serves as the base case for Step 3, and Lemma S.20 justifies Steps 2 and 3. Theorem S.22 shows that Step 4 works, which gives our main result for the order of the pole of the mean first passage time from each state $x \in \mathcal{B}^c$ to $\mathcal{B}$.

We will start with Sections S.5.1 and S.5.2, in which we describe in more detail the Big Theta notation used in this section and define some useful stopping times that will be used in our proof.

S.5.1 More on Big Theta notation In Section 4.2.1, we have defined orders for analytic functions using Big Theta notation. Here we give a few more definitions and remarks for inequalities involving the Big Theta notation, on how to compare the orders of analytic functions and on arithmetic for orders. These conventions streamline the proofs in the following subsections.

Definition S.16. *Given $\varepsilon_0 > 0$ and a function $f : (0, \varepsilon_0) \to \mathbb{R}_{>0}$, we say $f \le \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and a strictly positive $M_f \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,*

$$f(\varepsilon) \le M_f \varepsilon^k.$$

We say $f \ge \Theta(\varepsilon^k)$ if there exist $k \in \mathbb{Z}$ and a strictly positive $m_f \in \mathbb{R}_{>0}$ such that, for all $0 < \varepsilon < \varepsilon_0$,

$$f(\varepsilon) \ge m_f \varepsilon^k.$$

Remark S.17. Let $k, k_1, k_2 \in \mathbb{Z}$ and $k_1 \le k \le k_2$. If $f = \Theta(\varepsilon^k)$, then $f \le \Theta(\varepsilon^{k_1})$ and $f \ge \Theta(\varepsilon^{k_2})$.

Remark S.18. For functions f and g mapping $(0, \varepsilon_0)$ into $\mathbb{R}_{>0}$, we write $f = g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} = \Theta(\varepsilon^k)$, $f \le g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} \le \Theta(\varepsilon^k)$, $f \ge g \cdot \Theta(\varepsilon^k)$ if $\frac{f}{g} \ge \Theta(\varepsilon^k)$.

Lemma S.19. *Let $k_1, k_2 \in \mathbb{Z}$, $\varepsilon_0 > 0$ and $f, g : (0, \varepsilon_0) \to \mathbb{R}_{>0}$. If $f = \Theta(\varepsilon^{k_1})$ and $g = \Theta(\varepsilon^{k_2})$, then*

$$\frac{1}{f} = \Theta(\varepsilon^{-k_1}), \qquad f + g = \Theta(\varepsilon^{\min\{k_1,k_2\}}), \qquad f \cdot g = \Theta(\varepsilon^{k_1+k_2}),$$

$$\max\{f, g\} = \Theta(\varepsilon^{\min\{k_1,k_2\}}), \qquad \min\{f, g\} = \Theta(\varepsilon^{\max\{k_1,k_2\}}).$$

We leave the proof of Lemma S.19 to the reader.

S.5.2 Stopping times $\tau_n^{\varepsilon,N}$ For each graph $G^{(N)} = (V^{(N)}, E^{(N)})$, $N \in \{0, 1, \dots, M\}$, recall that $\{S(u) : u \in V^{(N)}\}$ is a partition of the state space $\mathcal{X}$. We define the series of stopping times $\{\tau_n^{\varepsilon,N}\}_{n=0}^{\infty}$, which captures times of transitions of X^ε between sets in the partition $\{S(u) : u \in V^{(N)}\}$ of $\mathcal{X}$. Formally, we let $\tau_0^{\varepsilon,N} = 0$, and for $n = 1, 2, \dots$, we successively define

$$\tau_n^{\varepsilon,N} = \inf\left\{t \ge \tau_{n-1}^{\varepsilon,N} : X^\varepsilon(t) \notin S(v_{n-1})\right\},$$

where v_{n-1} is the element in $V^{(N)}$ such that $X^\varepsilon(\tau_{n-1}^{\varepsilon,N}) \in S(v_{n-1})$.

S.5.3 Justification for Step 3 of the algorithm

Lemma S.20. *(i) For $N = 0$ in Step 3,*

(a) *for each $u \in V^{(0)} \setminus \{a\}$ and $x \in S(u)$, $\mathbb{E}_x[\tau_1^{\varepsilon,0}] = \Theta(\varepsilon^{-p_0(u)})$.*

(b) *for each $(u, v) \in E^{(0)}$ and $x \in S(u)$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,0}) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{uv}})$.*

(ii) For $N \in \{1, 2, \ldots, M\}$ in Step 3, let

$$k = \min\{\mathcal{K}_{uv} : u \in C_N, v \notin C_N \text{ and } (u, v) \in E^{(N-1)}\}.$$

(We note that k depends on N although we will not indicate that in the notation.)

(a) *For each $x \in S(c_N)$, $\mathbb{E}_x[\tau_1^{\varepsilon,N}] = \Theta(\varepsilon^{-p_0^x(c_N)})$ where $p_0^x(c_N) = p_0(c_N)$ and*

$$p_0(c_N) = \max\{p_0(u) : u \in C_N\} + k,$$

(b) *For each $x \in S(c_N)$ and $w \in V^{(N-1)} \setminus C_N$ such that $(u, w) \in E^{(N-1)}$ for some $u \in C_N$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(w)] = \Theta(\varepsilon^{\mathcal{K}^x_{c_N w}})$ where $\mathcal{K}^x_{c_N w} = \mathcal{K}_{c_N w}$ and*

$$\mathcal{K}_{c_N w} = \min\{\mathcal{K}_{uw} : u \in C_N \text{ and } (u, w) \in E^{(N-1)}\} - k,$$

(c) *For each $x \in S(w)$ where $w \in V^{(N-1)} \setminus C_N$ is such that $(w, v) \in E^{(N-1)}$ for some $v \in C_N$, $\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(c_N)] = \Theta(\varepsilon^{\mathcal{K}^x_{w c_N}})$ where $\mathcal{K}^x_{w c_N} = \mathcal{K}_{w c_N}$ and*

$$\mathcal{K}_{w c_N} = \min\{\mathcal{K}_{wv} : v \in C_N \text{ and } (w, v) \in E^{(N-1)}\}.$$

Proof. Our proof proceeds by induction. The base case ($N = 0$) is established in Section S.4.

For fixed $1 \le N \le M$, assume that *(i) (a)-(b) and (ii) (a)-(c)* hold with N replaced by $0, 1, \ldots, N-1$. We abbreviate $\tau_n^{\varepsilon,N-1}$ as τ_n^ε for $n = 0, 1, 2, \ldots$. Let

$$\delta_{out}(C_N) = \{(u, v) \in E^{(N-1)} : u \in C_N \text{ and } v \notin C_N\}$$

denote all out-going boundary edges of C_N so that

$$k = \min\{\mathcal{K}_{uv} : (u, v) \in \delta_{out}(C_N)\}.$$

First, consider the discrete time process $\{X^\varepsilon(\tau_n^\varepsilon)\}_{n=0}^\infty$, which is not necessarily a Markov process. We will derive a lower bound and an upper bound for

$$\mathbb{E}_x\left[\sum_{n=0}^\infty \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \le m \le n\}}\right], \tag{S.36}$$

which is the expected amount of time that $\{Y_n^\varepsilon = X^\varepsilon(\tau_n^\varepsilon)\}_{n=0}^\infty$ spends in $S(c_N)$ before exiting from there, when started from a fixed state $x \in S(c_N)$.

For the lower bound, let

$$\rho_1 = \max_{y \in S(c_N)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)],$$

the maximum over $y \in S(c_N)$ of the probability that for X^ε started at y, when X^ε exits $S(u_y)$, where $u_y \in C_N$ such that $y \in S(u_y)$, X^ε exits outside of $S(c_N)$. By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{uv}})$ for each u, v such that $(u, v) \in \delta_{out}(C_N)$ and $y \in S(u)$. Thus, using Lemma S.19, we have

$$\rho_1 = \max_{y \in S(c_N)} \sum_{\substack{(u,v)\in\delta_{out}(C_N):\\ y\in S(u)}} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\min\{\mathcal{K}_{uv}:(u,v)\in\delta_{out}(C_N)\}}) = \Theta(\varepsilon^k).$$

For $x \in S(c_N)$, let $\phi_n(x) = \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N)$ for $0 \le m \le n]$ for $n = 0, 1, 2, \ldots$. Then, $\phi_0(x) = 1$, $\phi_1(x) = \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(c_N)] \ge 1 - \rho_1$, and by the strong Markov property, for $n \ge 2$,

$$\begin{aligned}\phi_n(x) &= \sum_{y\in S(c_N)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \le m \le n-2; X^\varepsilon(\tau_{n-1}^\varepsilon) = y]\ \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(c_N)] \\ &\ge \phi_{n-1}(x)(1-\rho_1).\end{aligned}$$

Hence, $\phi_n(x) \ge (1-\rho_1)^n$ for $n = 0, 1, 2, \ldots$. Then, for $x \in S(c_N)$,

$$\text{(S.37)} \qquad \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in S(c_N) \text{ for } 0\le m\le n\}}\right] = \sum_{n=0}^{\infty}\phi_n(x) \ge \sum_{n=0}^{\infty}(1-\rho_1)^n = \frac{1}{\rho_1} = \Theta(\varepsilon^{-k}).$$

and so (S.36) is bounded below by $\Theta(\varepsilon^{-k})$.

For the upper bound, let $w_0 \in C_N$ be such that

$$\min\{\mathcal{K}_{w_0 v} : (w_0, v) \in \delta_{out}(C_N)\} = \min\{\mathcal{K}_{uv} : (u, v) \in \delta_{out}(C_N)\} = k.$$

Since the order of the probability $\mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)]$ might equal $k' > k$ for some $w \ne w_0 \in C_N$ and $x \in S(w)$, such a smaller order probability of directly exiting $S(c_N)$ from $S(w)$ makes it seem possible that (S.36) could be $\Theta(\varepsilon^{-k'})$ for some $k' > k$. Indeed, using a similar approach to the one we used for the lower bound, we can show that (S.36) is bounded above by $\Theta(\varepsilon^{\max\{\mathcal{K}_{uv}:(u,v)\in\delta_{out}(C_N)\}}) \ge \Theta(\varepsilon^k)$. However, we would like a more stringent upper bound. To achieve this, we will show below that from $S(w)$, X^ε can exit $S(c_N)$ at least as quickly by means of a transition from $S(w)$ to $S(w_0)$ via the r-connected set and then from $S(w_0)$ to $V^{(N-1)} \setminus S(c_N)$.

Let $\zeta_0^\varepsilon = 0$, and for $n = 1, 2, \ldots$, we successively define

$$\eta_{n-1}^\varepsilon = \inf\left\{t \ge \zeta_{n-1}^\varepsilon : X^\varepsilon(t) \notin S(v) \text{ where } v \in V^{(N-1)} \text{ and } X^\varepsilon(\zeta_{n-1}^\varepsilon) \in S(v)\right\},$$

$$\zeta_n^\varepsilon = \inf\left\{t \ge \eta_{n-1}^\varepsilon : X^\varepsilon(t) \in S(w_0) \text{ or } X^\varepsilon(t) \notin S(c_N)\right\}.$$

Note that $\{\zeta_n^\varepsilon\}_{n=0}^\infty$ and $\{\eta_n^\varepsilon\}_{n=0}^\infty$ depend on N. Let

$$\rho_2 = \min_{y\in S(w_0)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w_0)] = \min_{y\in S(w_0)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(c_N)],$$

By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{w_0 v}})$ for each v such that $(w_0, v) \in \delta_{out}(C_N)$ and $y \in S(w_0)$, and so

$$\rho_2 \geq \min_{y \in S(w_0)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)] = \Theta(\varepsilon^{\min\{\mathcal{K}_{w_0 v}:(w_0,v)\in\delta_{out}(C_N)\}}) = \Theta(\varepsilon^k), \tag{S.38}$$

where the inequality holds since starting from any $y \in S(w_0)$, if $X^\varepsilon(\tau_1^\varepsilon) \notin S(c_N)$, X^ε exits outside of $S(c_N)$ after leaving $S(w_0)$ and so $X^\varepsilon(\zeta_1^\varepsilon) \notin S(w_0)$.

For $x \in S(c_N)$, $\mathbb{P}_x$-a.s., the sum

$$\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in S(c_N) \text{ for } 0\leq m<n; X^\varepsilon(\tau_n^\varepsilon)\in S(w_0)\}} \tag{S.39}$$

counts the number of distinct visits to $S(w_0)$, including the initial start there if $x \in S(w_0)$, before X^ε escapes from $S(c_N)$. By the definition of the $\{\zeta_n^\varepsilon\}_{n=0}^\infty$, $\mathbb{P}_x$-a.s., the sum

$$\mathbb{1}_{\{X^\varepsilon(0)\in S(w_0)\}} + \sum_{n=1}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w_0) \text{ for } 1\leq m\leq n\}}$$

counts the same quantity. Thus, for $x \in S(w_0)$, using the strong Markov property and (S.38),

$$\begin{aligned}
\psi(x) &:= \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in S(c_N) \text{ for } 0\leq m<n; X^\varepsilon(\tau_n^\varepsilon)\in S(w_0)\}}\right] \\
&= 1 + \mathbb{E}_x\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w_0) \text{ for } 1\leq m\leq n\}}\right] \\
&= 1 + \mathbb{E}_x\left[\mathbb{1}_{\{X^\varepsilon(\zeta_1^\varepsilon)\in S(w_0)\}} \mathbb{E}_{X^\varepsilon(\zeta_1^\varepsilon)}\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w_0) \text{ for } 0\leq m\leq n\}}\right]\right] \\
&\leq 1 + (1-\rho_2) \max_{y\in S(w_0)} \psi(y).
\end{aligned} \tag{S.40}$$

Note that $\max_{y\in S(w_0)} \psi(y) < \infty$ because the state space is finite and X^ε is positive recurrent. Hence, by (S.40), $\max_{y\in S(w_0)} \psi(y) \leq \frac{1}{\rho_2}$. Then, for $x \in S(c_N) \setminus S(w_0)$, by the strong Markov property,

$$\psi(x) \leq \mathbb{P}_x[X^\varepsilon(\zeta_1^\varepsilon) \in S(w_0)\}] \max_{y\in S(w_0)} \psi(y) \leq \frac{1}{\rho_2}.$$

Thus, for any $x \in S(c_N)$,

$$\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in S(c_N) \text{ for } 0\leq m<n; X^\varepsilon(\tau_n^\varepsilon)\in S(w_0)\}}\right] \leq \frac{1}{\rho_2} = \Theta(\varepsilon^{-k}). \tag{S.41}$$

Let $w_1, w_2 \in C_N$ be such that $(w_1, w_2) \in E^{(N-1)}$ and $\mathcal{K}_{w_1 w_2} = 0$. By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)] = \Theta(1)$ for all $y \in S(w_1)$. Then, for $x \in S(c_N)$,

$$\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in S(c_N) \text{ for } 0\leq m<n; X^\varepsilon(\tau_n^\varepsilon)\in S(w_2)\}}\right]$$

$$\geq \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1), X^\varepsilon(\tau_{n+1}) \in S(w_2)]$$

$$= \sum_{n=0}^{\infty} \sum_{y \in S(w_1)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)]$$

$$\geq \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1)] \cdot \min_{y \in S(w_1)} \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_2)]$$

$$= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_1)\}} \right] \cdot \Theta(1). \tag{S.42}$$

where the first equality holds from the strong Markov property of X^ε. Since C_N is an r-connected set, we can start from the node w_1 and the order inequality (S.42) can be passed from node to node in C_N and back to the node w_1 (w_0 is included in the path) so that we will actually have equality in (S.42) and for all $v \in C_N$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(v)\}} \right]$$

$$= \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w_0)\}} \right] \cdot \Theta(1). \tag{S.43}$$

Therefore, combining (S.41) and (S.43), and since there are only finitely many nodes in C_N, we can obtain by summing over $v \in C_N$ that (S.36) is bounded above by $\Theta(\varepsilon^{-k})$. Combining with (S.37), we have that, for $x \in S(c_N)$, (S.36) is $\Theta(\varepsilon^{-k})$. Moreover, by (S.43), for each $x \in S(c_N)$ and each $v \in C_N$,

$$\mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(v)\}} \right] = \Theta(\varepsilon^{-k}). \tag{S.44}$$

To prove *(ii) (a)*, fix $x \in S(c_N)$. By the induction hypothesis, $\mathbb{E}_y[\tau_1^\varepsilon] = \Theta(\varepsilon^{-p_0(u)})$ for each $y \in S(u)$ where $u \in C_N$. Thus, the expected sojourn time in $S(c_N)$ is

$$\mathbb{E}_x[\tau_1^{\varepsilon,N}] = \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m \leq n\}} \cdot (\tau_{n+1}^\varepsilon - \tau_n^\varepsilon) \right]$$

$$= \sum_{n=0}^{\infty} \sum_{u \in C_N} \sum_{y \in S(u)} \mathbb{E}_x[\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) = y\}}] \cdot \mathbb{E}_y[\tau_1^\varepsilon - \tau_0^\varepsilon]$$

$$= \sum_{u \in C_N} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \leq m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(u)\}} \right] \cdot \Theta(\varepsilon^{-p_0(u)})$$

$$= \sum_{u \in C_N} \Theta(\varepsilon^{-k-p_0(u)}) = \Theta(\varepsilon^{-k-\max\{p_0(u): u \in C_N\}}) = \Theta(\varepsilon^{-p_0(c_N)}).$$

where the first equality holds from the strong Markov property of X^ε, we used (S.44) for the third equality, and we used Lemma S.19 for the fourth equality.

To prove *(ii) (b)*, fix $x \in S(c_N)$ and $w \in V^{(N-1)} \setminus C_N$. By the induction hypothesis, $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)] = \Theta(\varepsilon^{\mathcal{K}_{uw}})$ for each $y \in S(u)$ where $u \in C_N$. Thus, starting from x, the probability of exiting $S(c_N)$ by means of a transition from a state in $S(c_N)$ to a state in $S(w)$ is given by

$$\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(w)] = \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \le m \le n; X^\varepsilon(\tau_{n+1}^\varepsilon) \in S(w)]$$

$$= \sum_{n=0}^{\infty} \sum_{u \in C_N} \sum_{y \in S(u)} \mathbb{E}_x[\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \le m < n; X^\varepsilon(\tau_n^\varepsilon) = y\}}] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)]$$

$$= \sum_{\substack{u \in C_N: \\ (u,w) \in E^{(N-1)}}} \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \in S(c_N) \text{ for } 0 \le m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(u)\}}\right] \cdot \Theta(\varepsilon^{\mathcal{K}_{uw}})$$

$$= \sum_{\substack{u \in C_N: \\ (u,w) \in E^{(N-1)}}} \Theta(\varepsilon^{-k+\mathcal{K}_{uw}}) = \Theta(\varepsilon^{-k+\min\{\mathcal{K}_{uw} : u \in C_N \text{ and } (u,w) \in E^{(N-1)}\}}) = \Theta(\varepsilon^{\mathcal{K}_{c_N w}}),$$

where we used (S.44) for the third equality, and for the second equality, we used the fact that there must be an edge in $E^{(N-1)}$ between u and w if $\mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w)] > 0$ for some and hence all $y \in S(u)$.

To prove *(ii) (c)*, fix $x \in S(w)$ where $w \in V^{(N-1)} \setminus C_N$. By the induction hypothesis, $\mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(\varepsilon^{\mathcal{K}_{wv}})$ for each $v \in C_N$. Thus, starting from x, the probability of entering $S(c_N)$ by means of a transition from a state in $S(w)$ to a state in $S(c_N)$ is

$$\mathbb{P}_x[X^\varepsilon(\tau_1^{\varepsilon,N}) \in S(c_N)] = \sum_{v \in C_N} \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(v)]$$

$$= \sum_{\substack{v \in C_N: \\ (w,v) \in E^{(N-1)}}} \Theta(\varepsilon^{\mathcal{K}_{wv}}) = \Theta(\varepsilon^{\min\{\mathcal{K}_{wv} : v \in C_N \text{ and } (w,v) \in E^{(N-1)}\}}) = \Theta(\varepsilon^{\mathcal{K}_{w c_N}}).$$

■

S.5.4 Justification for Step 4 of the algorithm

Lemma S.21. *Fix $w \in V^{(M,0)} \setminus \{a\}$. Let $\tau_n^\varepsilon = \tau_n^{\varepsilon,M}$, for $n = 0, 1, 2, \ldots$, as defined in Section S.5.2. Then, starting from $x \in S(w)$, the expected number of distinct visits to $S(w)$, including the initial start there, before X^ε enters $S(a)$ is*

$$\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon) \notin S(a) \text{ for } 0 \le m < n; X^\varepsilon(\tau_n^\varepsilon) \in S(w)\}}\right] = \Theta(1).$$

Proof. Let $\zeta_0^\varepsilon = 0$, and for $n = 1, 2, \ldots$, successively define

$$\eta_{n-1}^\varepsilon = \inf\left\{t \ge \zeta_{n-1}^\varepsilon : X^\varepsilon(t) \notin S(v) \text{ where } v \in V^{(M,0)} \text{ and } X^\varepsilon(\zeta_{n-1}^\varepsilon) \in S(v)\right\},$$

$$\zeta_n^\varepsilon = \inf\left\{t \ge \eta_{n-1}^\varepsilon : X^\varepsilon(t) \in S(w) \cup S(a)\right\}.$$

Note that for $x \in S(w)$, $\mathbb{P}_x$-a.s.,

$$\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\notin S(a) \text{ for } 0\le m<n; X^\varepsilon(\tau_n^\varepsilon)\in S(w)\}} = \sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w) \text{ for } 0\le m\le n\}}, \tag{S.45}$$

since they both count the number of distinct visits to $S(w)$, including the initial start there, before X^ε enters $S(a)$.

Recall from Section S.4 that for each $u \in V^{(M,0)} \setminus \{a\}$, there is an r-path from u to a. Let such an r-path from w to a be $w \to w_1 \ldots \to w_d \to a$ where $w, w_1, \ldots, w_d, a$ are distinct. By definition, an edge $(u,v) \in V^{(M,0)}$ is an r-edge implies that $\mathbb{P}_z[X^\varepsilon(\tau_1^\varepsilon) \in S(v)] = \Theta(1)$ for all $z \in S(u)$. Thus, for any $y \in S(w)$, using the strong Markov property of X^ε, we have

$$\begin{aligned}
\Theta(1) = 1 &\ge \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)] = \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \in S(a)] \\
&\ge \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \ldots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d), X^\varepsilon(\tau_{d+1}^\varepsilon) \in S(a)] \\
&= \sum_{z\in S(w_d)} \mathbb{P}_x[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \ldots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d), X^\varepsilon(\tau_d^\varepsilon) = z] \cdot \mathbb{P}_z[X^\varepsilon(\tau_1^\varepsilon) \in S(a)] \\
&= \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(w_1), \ldots, X^\varepsilon(\tau_d^\varepsilon) \in S(w_d)] \cdot \Theta(1) = \ldots = \Theta(1) \cdot \ldots \cdot \Theta(1) = \Theta(1).
\end{aligned}$$

Using a similar approach to that used in Section S.5.3, we can show that for $x \in S(w)$,

$$\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w) \text{ for } 0\le m\le n\}}\right] \ge \frac{1}{\max_{y\in S(w)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)]} = \Theta(1),$$

and

$$\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\zeta_m^\varepsilon)\in S(w) \text{ for } 0\le m\le n\}}\right] \le \frac{1}{\min_{y\in S(w)} \mathbb{P}_y[X^\varepsilon(\zeta_1^\varepsilon) \notin S(w)]} = \Theta(1).$$

Combining these inequalities with (S.45) yields the desired result. ■

Theorem S.22. *Let $\tau_{\mathcal{B}}^\varepsilon = \inf\{t \ge 0 : X^\varepsilon(t) \in \mathcal{B}\}$ be the first passage time to $\mathcal{B}$ for X^ε. For each $N = 1, \ldots, |V^{(M)}| - 1$ and $x \in S(v_N)$, we have*

$$\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] = \Theta(\varepsilon^{-p(v_N)}). \tag{S.46}$$

Proof. It suffices by iteration to prove that for each fixed $1 \le N \le |V^{(M)}| - 1$, if

$$\mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon] = \Theta(\varepsilon^{-p(v_k)}) \text{ for all } y \in S(v_k) \text{ and } 1 \le k \le N-1, \tag{S.47}$$

then (S.46) holds for all $x \in S(v_N)$. By convention, (S.47) holds automatically for $N = 1$.

For the iteration step, fix $1 \le N \le |V^{(M)}| - 1$, and assume that (S.47) holds. If $N = 1$, let $A = V^{(M,0)} \setminus \{a\}$, and if $N > 1$, let $A = V^{(M,0)} \setminus \{v_1, \ldots, v_{N-1}, a\}$. Recall that for $w \in A$ and $1 \le k \le N-1$, we have $w \in V^{(M,k)}$, and so by (S.35),

$$p_k(w) = \begin{cases} \max\{p_{k-1}(w), p_{k-1}(v_k) - \mathcal{K}_{wv_k}\} & \text{for } (w, v_k) \in E^{(M,k-1)}, \\ p_{k-1}(w), & \text{for } (w, v_k) \notin E^{(M,k-1)}. \end{cases} \tag{S.48}$$

Note that $(w, v_k) \in E^{(M,k-1)}$ if and only if $(w, v_k) \in E^{(M,0)}$. Since $p_{k-1}(v_k) =: p(v_k)$ for $1 \le k \le N-1$, by iterating (S.48), we can obtain

$$p_{N-1}(w) = \max\{p_0(w), \max\{p(v_k) - \mathcal{K}_{wv_k} : 1 \le k \le N-1 \text{ and } (w, v_k) \in E^{(M,0)}\}\}, \tag{S.49}$$

where we make the convention that a maximum over an empty set is $-\infty$. In particular, since $v_N \in A$, we have

$$p(v_N) := p_{N-1}(v_N) = \max\{p_0(v_N), \max\{p(v_k) - \mathcal{K}_{v_N v_k} : 1 \le k \le N-1 \text{ and } (v_N, v_k) \in E^{(M,0)}\}\}. \tag{S.50}$$

Fix $x \in S(v_N)$. We will derive a lower bound and an upper bound for $\mathbb{E}_x[\tau^\varepsilon_{\mathcal{B}}]$. For the lower bound, let $\tau^\varepsilon = \inf\{t \ge 0 : X^\varepsilon(t) \notin S(v_N)\}$. Recall that $\mathbb{P}_x$-a.s., $\tau^\varepsilon = \tau_1^{\varepsilon,M}$ as defined in Section S.5.2. By Lemma S.20, for each $y \in S(w)$ where $w \in V^{(M,0)} = V^{(M)}$ is such that $(v_N, w) \in E^{(M,0)} = E^{(M)}$, $\mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] = \Theta(\varepsilon^{\mathcal{K}_{v_N w}})$ and $\mathbb{E}_x[\tau^\varepsilon] = \Theta(\varepsilon^{-p_0(v_N)})$. By first step analysis,

$$\begin{aligned}
\mathbb{E}_x[\tau^\varepsilon_{\mathcal{B}}] &= \mathbb{E}_x[\tau^\varepsilon] + \sum_{(v_N,w)\in E^{(M,0)}} \sum_{y\in S(w)} \mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] \cdot \mathbb{E}_y[\tau^\varepsilon_{\mathcal{B}}] \\
&\ge \mathbb{E}_x[\tau^\varepsilon] + \sum_{\substack{1\le k\le N-1:\\ (v_N,v_k)\in E^{(M,0)}}} \sum_{y\in S(v_k)} \mathbb{P}_x[X^\varepsilon(\tau^\varepsilon) = y] \cdot \mathbb{E}_y[\tau^\varepsilon_{\mathcal{B}}] \\
&= \Theta(\varepsilon^{-p_0(v_N)}) + \sum_{\substack{1\le k\le N-1:\\ (v_N,v_k)\in E^{(M,0)}}} \Theta(\varepsilon^{\mathcal{K}_{v_N v_k}}) \cdot \Theta(\varepsilon^{-p(v_k)}) = \Theta(\varepsilon^{-p_{N-1}(v_N)}). \qquad \text{(S.51)}
\end{aligned}$$

where we used (S.47) in the second last equality, and used Lemma S.19 and (S.50) for the last equality.

For the upper bound, let $\eta^\varepsilon = \inf\{t \ge 0 : X^\varepsilon(t) \notin \bigcup_{u\in A} S(u)\}$. Let $\tau_n^\varepsilon = \tau_n^{\varepsilon,M}$, for $n = 0, 1, 2, \dots$, as defined in Section S.5.2. Then, using Lemma S.21 and the strong Markov property, for $w \in A \subset V^{(M,0)} \setminus \{a\}$,

$$\begin{aligned}
&\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in\bigcup_{u\in A} S(u) \text{ for } 0\le m<n;\, X^\varepsilon(\tau_n^\varepsilon)\in S(w)\}}\right] \\
\text{(S.52)} \le\; & \mathbb{P}_x[X^\varepsilon(\zeta^\varepsilon) \in S(w)] \max_{y\in S(w)} \mathbb{E}_y\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\notin S(a) \text{ for } 0\le m<n;\, X^\varepsilon(\tau_n^\varepsilon)\in S(w)\}}\right] \le \Theta(1).
\end{aligned}$$

where $\zeta^\varepsilon = \inf\{t \ge 0 : X^\varepsilon(t) \in S(w)\}$.

For $1 \le k \le N-1$ such that there exists $w \in A$ where $(w, v_k) \in E^{(M,0)}$,

$$\begin{aligned}
\mathbb{P}_x[X^\varepsilon(\eta^\varepsilon) \in S(v_k)] &= \sum_{n=0}^{\infty} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u\in A} S(u) \text{ for } 0 \le m \le n;\, X^\varepsilon(\tau_{n+1}^\varepsilon) \in S(v_k)] \\
&= \sum_{n=0}^{\infty} \sum_{w\in A} \sum_{y\in S(w)} \mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon) \in \bigcup_{u\in A} S(u) \text{ for } 0 \le m < n;\, X^\varepsilon(\tau_n^\varepsilon) = y] \cdot \mathbb{P}_y[X^\varepsilon(\tau_1^\varepsilon) \in S(v_k)]
\end{aligned}$$

$$= \sum_{\substack{w\in A:\\(w,v_k)\in E^{(M,0)}}} \mathbb{E}_x\left[\sum_{n=0}^{\infty}\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in\bigcup_{u\in A}S(u)\text{ for }0\le m<n;X^\varepsilon(\tau_n^\varepsilon)\in S(w)\}}\right]\cdot\Theta(\varepsilon^{\mathcal{K}_{wv_k}})$$

$$\le \sum_{\substack{w\in A:\\(w,v_k)\in E^{(M,0)}}} \Theta(1)\cdot\Theta(\varepsilon^{\mathcal{K}_{wv_k}}), \tag{S.53}$$

where the second equality holds from strong Markov property of X^ε, the third equality uses Lemma S.20, and we used (S.52) for the last inequality. Using Lemma S.20 and (S.52),

$$\mathbb{E}_x[\eta^\varepsilon] = \mathbb{E}_x\left[\sum_{n=0}^{\infty}\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in\bigcup_{u\in A}S(u)\text{ for }0\le m\le n\}}\cdot(\tau_{n+1}^\varepsilon-\tau_n^\varepsilon)\right]$$

$$= \sum_{n=0}^{\infty}\sum_{w\in A}\sum_{y\in S(w)}\mathbb{P}_x[X^\varepsilon(\tau_m^\varepsilon)\in\bigcup_{u\in A}S(u)\text{ for }0\le m<n;X^\varepsilon(\tau_n^\varepsilon)=y]\cdot\mathbb{E}_y[\tau_1^\varepsilon-\tau_0^\varepsilon]$$

$$= \sum_{w\in A}\mathbb{E}_x\left[\sum_{n=0}^{\infty}\mathbb{1}_{\{X^\varepsilon(\tau_m^\varepsilon)\in\bigcup_{u\in A}S(u)\text{ for }0\le m<n;X^\varepsilon(\tau_n^\varepsilon)\in S(w)\}}\right]\cdot\Theta(\varepsilon^{-p_0(w)})$$

$$\le \sum_{w\in A}\Theta(1)\cdot\Theta(\varepsilon^{-p_0(w)})\le\Theta(\varepsilon^{-p_{N-1}(v_N)}), \tag{S.54}$$

where we have used (S.49) and (S.34) to conclude that $p_0(w)\le p_{N-1}(w)\le p_{N-1}(v_N)$ for all $w\in A$. Therefore, using first step analysis, we have

$$\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon] = \mathbb{E}_x[\eta^\varepsilon] + \sum_{1\le k\le N-1}\sum_{y\in S(v_k)}\mathbb{P}_x[X^\varepsilon(\eta^\varepsilon)=y]\cdot\mathbb{E}_y[\tau_{\mathcal{B}}^\varepsilon]$$

$$= \mathbb{E}_x[\eta^\varepsilon] + \sum_{1\le k\le N-1}\mathbb{P}_x[X^\varepsilon(\eta^\varepsilon)\in S(v_k)]\cdot\Theta(\varepsilon^{-p(v_k)})$$

$$\le \Theta(\varepsilon^{-p_{N-1}(v_N)}) + \sum_{\substack{w\in A,\\1\le k\le N-1:\\(w,v_k)\in E^{(M,0)}}}\Theta(\varepsilon^{\mathcal{K}_{wv_k}})\cdot\Theta(\varepsilon^{-p(v_k)})$$

$$\le \Theta(\varepsilon^{-p_{N-1}(v_N)}) + \Theta(\varepsilon^{-\max\{p_{N-1}(w):w\in A\}}) = \Theta(\varepsilon^{-p_{N-1}(v_N)}), \tag{S.55}$$

where we used (S.47) for the second equality, (S.53) and (S.54) for the first inequality, and (S.49) and Lemma S.19 for the second inequality.

By (S.51) and (S.55), we conclude that $\mathbb{E}_x[\tau_{\mathcal{B}}^\varepsilon]=\Theta(\varepsilon^{-p_{N-1}(v_N)})=\Theta(\varepsilon^{-p(v_N)})$. ■

S.6 Application of the algorithm to the 2D, 3D and 4D models The algorithm is described in Section 4.2.1, and it finds the order of the pole of the mean first passage time to $\emptyset\ne\mathcal{B}\subset\mathcal{X}$ from each state in $\mathcal{B}^c$. In this section, we will apply the algorithm to the 2D, 3D and 4D models and find the order of the poles of the mean first passage times of interest to the fully repressed state and the fully active state (Figure S.1 – S.5). For each figure, the "Input" panel shows the order of each of the non-zero off-diagonal entries in $Q(\varepsilon)$ and the set $\mathcal{B}$ which contains a single state, which is either the fully repressed state or the fully active state. The orders of these non-zero entries in $Q(\varepsilon)$ are represented by colored arrows in the graph (red for order 0 and blue for order 1). Step 1 transforms the orders in the infinitesimal generator $Q(\varepsilon)$ into orders for the transition matrix $P(\varepsilon)$ and the exponential parameters $q(\varepsilon)$ to give an equivalent construction for the continuous time Markov chain. The orders of the non-zero entries in $P(\varepsilon)$ are given by $\mathcal{K}$ and represented by colored arrows in the graph, and the number in the circle at a state $x \in \mathcal{B}^c$ is the order of the pole $p(x)$ of $\frac{1}{q_x(\varepsilon)}$ (the mean sojourn time at the state x). In Step 2, the set $\mathcal{B}$ contains only one state and is just relabeled as the node a. All transitions from a to $\mathcal{B}^c$ are then removed. While the Input, Step 1 and Step 2 are universal across all the figures in this section, we explain the Step 3, Step 4 and Output panels separately for each application below since they are more distinct.

2D model (from the fully active state to the fully repressed state): see Figure S.1. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(\mathrm{D_{tot}}, 0)^T\}$. Step 3 for the 2D model involves only one iteration, where the collection of all nodes except the node a and the origin 0 (called an r-connected set C) is condensed to a single node c, and the order of the pole at c is $p(c) = \max_{u\in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1+0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c0} = \min\{\mathcal{K}_{u0} : u \in C \text{ and } (u, 0) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1-0 = 1$, $\mathcal{K}_{ca} = \min\{\mathcal{K}_{ua} : u \in C \text{ and } (u, a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0-0 = 0$, and $\mathcal{K}_{0c} = \min\{\mathcal{K}_{0v} : v \in C \text{ and } (0, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p, which is c in our case. At any node other than a that is connected to c (i.e., the origin 0), the value of p is updated to $p(0) = \max\{p(0), p(c) - \mathcal{K}_{0c}\} = \max\{0, 1 - 0\} = 1$, and then any edges leading to or from c are removed. In the second iteration, of the remaining nodes, we fix the node with the largest value of p, which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

2D model (from the fully repressed state to the fully active state) Because of the symmetry in the input graph in Figure S.1, the orders of the poles of the mean first passage times to the fully repressed state can be obtained in the same way as above.

3D model (from the fully active state to the fully repressed state): see Figure S.2. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(\mathrm{D_{tot}}, 0, 0)^T\}$. A state represents $(n_{\mathrm{D^R_{12}}}, n_{\mathrm{D^A}}, n_{\mathrm{D^R_1}})^T$. Step 3 involves only one iteration, where the collection of all nodes except for $(0, 0, 0)^T$, $(0, 0, \mathrm{D_{tot}})^T$, $(1, 0, \mathrm{D_{tot}}-1)^T$, $(2, 0, \mathrm{D_{tot}}-2)^T$,. . ., $(\mathrm{D_{tot}}-2, 0, 2)^T$, and $(\mathrm{D_{tot}}-1, 0, 1)^T$ (called an r-connected set C) is condensed to a single node c. The order of the pole of the sojourn time at C is $p(c) = \max_{u\in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c,(0,0,0)^T} = \min\{\mathcal{K}_{u,(0,0,0)^T} : u \in C \text{ and } (u, (0, 0, 0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{(0,0,0)^T,c} = \min\{\mathcal{K}_{(0,0,0)^T,v} : v \in C \text{ and } ((0, 0, 0)^T, v) \in E\} = 0$, $\mathcal{K}_{c,(0,0,\mathrm{D_{tot}})^T} = \min\{\mathcal{K}_{u,(0,0,\mathrm{D_{tot}})^T} : u \in C \text{ and } (u, (0, 0, \mathrm{D_{tot}})^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0 - 0 = 0$, $\mathcal{K}_{(0,0,\mathrm{D_{tot}})^T,c} = \min\{\mathcal{K}_{(0,0,\mathrm{D_{tot}})^T,v} : v \in C \text{ and } ((0, 0, \mathrm{D_{tot}})^T, v) \in E\} = 1$, . . ., and $\mathcal{K}_{c,(\mathrm{D_{tot}}-1,0,1)^T} = \min\{\mathcal{K}_{u,(\mathrm{D_{tot}}-1,0,1)^T} : u \in C \text{ and } (u, (\mathrm{D_{tot}} - 1, 0, 1)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u, v) \in E\} = 0 - 0 = 0$, $\mathcal{K}_{(\mathrm{D_{tot}}-1,0,1)^T,c} = \min\{\mathcal{K}_{(\mathrm{D_{tot}}-1,0,1)^T,v} : v \in C \text{ and } ((\mathrm{D_{tot}} - 1, 0, 1)^T, v) \in E\} = 1$. Step 4 involves $(\mathrm{D_{tot}} + 2)$ iterations. In the first iteration, we fix the node with the largest value of p, which is c in our case. At any node u other than a that is connected to c, the value of p is updated according to the formula $p(u) = \max\{p(u), p(c) - \mathcal{K}_{uc}\}$, and then any edges leading to or from c are removed. In the second iteration, the node $(0,0,0)^T$ has the largest value of p among the remaining nodes, and thus is fixed. There is no other nodes connected to $(0,0,0)^T$ at this point, so we move to the next iteration. In the third iteration, the node $(0,0,\mathrm{D_{tot}})^T$ is fixed. The node $(1,0,\mathrm{D_{tot}}-1)^T$ is connected to it, and thus the $p((1,0,\mathrm{D_{tot}}-1)^T)$ is updated to be $\max\{p((1,0,\mathrm{D_{tot}}-1)^T), p((0,0,\mathrm{D_{tot}})^T) - \mathcal{K}_{(1,0,\mathrm{D_{tot}}-1)^T,(0,0,\mathrm{D_{tot}})^T}\} = 0$. Then, any edges leading to or from $(0,0,\mathrm{D_{tot}})^T$ are removed. The remaining iterations will be similar to the third one. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

3D model (from the fully repressed state to the fully active state): see Figure S.3. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(0,\mathrm{D_{tot}},0)^T\}$. Step 3 involves two iterations. In the first iteration, the collection of nodes consisting of $(\mathrm{D_{tot}}-1,0,1)^T$ and $(\mathrm{D_{tot}},0,0)^T$ (called an r-connected set C_1) is condensed into a single node c_1. The order of the pole of the sojourn time in C_1 is $p(c_1) = \max_{u\in C_1} p(u) + \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u,v) \in E\} = 1 + 1 = 2$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, $\mathcal{K}_{c_1,(\mathrm{D_{tot}}-2,0,2)^T} = \min\{\mathcal{K}_{u,(\mathrm{D_{tot}}-2,0,2)^T} : u \in C_1 \text{ and } (u,(\mathrm{D_{tot}}-2,0,2)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u,v) \in E\} = 1 - 1 = 0$, $\mathcal{K}_{(\mathrm{D_{tot}}-2,0,2)^T,c_1} = \min\{\mathcal{K}_{(\mathrm{D_{tot}}-2,0,2)^T,v} : v \in C_1 \text{ and } ((\mathrm{D_{tot}}-2,0,2)^T, v) \in E\} = 0$, $\mathcal{K}_{c_1,(\mathrm{D_{tot}}-1,0,0)^T} = \min\{\mathcal{K}_{u,(\mathrm{D_{tot}}-1,0,0)^T} : u \in C_1 \text{ and } (u,(\mathrm{D_{tot}}-1,0,0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u,v) \in E\} = 1 - 1 = 0$ and $\mathcal{K}_{(\mathrm{D_{tot}}-1,0,0)^T,c_1} = \min\{\mathcal{K}_{(\mathrm{D_{tot}}-1,0,0)^T,v} : v \in C_1 \text{ and } ((\mathrm{D_{tot}}-1,0,0)^T, v) \in E\} = 0$. In the second iteration of Step 3, the collection of all nodes except for $(0,0,0)^T$ and a (called an r-connected set C_2) is condensed to a single node c_2. The order of the pole of the sojourn time in C_2 is $p(c_2) = \max_{u\in C_2} p(u) + \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 2 + 0 = 2$, where E denotes the edge set of the graph in Step 3 before the second iteration. Moreover, $\mathcal{K}_{c_2,(0,0,0)^T} = \min\{\mathcal{K}_{u,(0,0,0)^T} : u \in C_2 \text{ and } (u,(0,0,0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{c_2,a} = \min\{\mathcal{K}_{ua} : u \in C_2 \text{ and } (u,a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 0 - 0 = 0$, and $\mathcal{K}_{(0,0,0)^T,c_2} = \min\{\mathcal{K}_{(0,0,0)^T,v} : v \in C_2 \text{ and } ((0,0,0)^T, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p, which is c_2 in our case. At any node other than a that is connected to c_2 (i.e., the origin $(0,0,0)^T$), the value of p is updated to $p((0,0,0)^T) = \max\{p((0,0,0)^T), p(c_2) - \mathcal{K}_{(0,0,0)^T,c_2}\} = \max\{0, 2-0\} = 2$, and then any edges leading to or from c_2 are removed. In the second iteration, among the remaining nodes, we fix the node with the largest value of p, which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

4D model (from the fully active state to the fully repressed state): see Figure S.4. We illustrate how to use the algorithm for the 4D model when $\mathrm{D_{tot}} = 2$; for larger $\mathrm{D_{tot}}$, the methodology will be the same. A state represents $(n_{\mathrm{D}^{\mathrm{R}}_{12}}, n_{\mathrm{D}^{\mathrm{A}}}, n_{\mathrm{D}^{\mathrm{R}}_{1}}, n_{\mathrm{D}^{\mathrm{R}}_{2}})^T$. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(2,0,0,0)^T\}$. Step 3 involves only one iteration, where the collection of all nodes except for $(0,0,0,0)^T$, $(0,0,2,0)^T$, $(0,0,1,1)^T$, $(0,0,0,2)^T$, $(1,0,1,0)^T$ and $(1,0,0,1)^T$ (called an r-connected set C) is condensed to a single node c. The order of the pole of the sojourn time in C is $p(c) =$

$\max_{u\in C} p(u) + \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u,v) \in E\} = 1 + 0 = 1$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, the value of $\mathcal{K}$ for edges between c and an original node w that is not in C are defined according to the formula $\mathcal{K}_{c,w} = \min\{\mathcal{K}_{uw} : u \in C \text{ and } (u,w) \in E\} - \min\{\mathcal{K}_{uv} : u \in C, v \notin C \text{ and } (u,v) \in E\}$, $\mathcal{K}_{w,c} = \min\{\mathcal{K}_{w,v} : v \in C \text{ and } (w,v) \in E\}$. Step 4 involves seven iterations. In the first iteration, we fix the node with the largest value of p, which is c in our case. At any node u other than a that is connected to c, the value of p is updated according to the formula $p(u) = \max\{p(u), p(c) - \mathcal{K}_{uc}\}$, and then any edges leading to or from c are removed. In the second iteration of Step 4, the node $(0,0,0,0)^T$ has the largest value of p among the remaining nodes, and then is fixed. There are no other nodes connected to $(0,0,0,0)^T$ at this point, so we move to the next iteration. In the third iteration, the node $(0,0,2,0)^T$ is fixed. The node $(1,0,1,0)^T$ is connected to it, and thus $p((1,0,1,0)^T)$ is updated to be $\max\{p((1,0,1,0)^T), p((0,0,2,0)^T) - \mathcal{K}_{(1,0,1,0)^T,(0,0,2,0)^T}\} = 0$. Then, any edges leading to or from $(0,0,2,0)^T$ are removed. The remaining iterations will be similar to the third one. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

4D model (from the fully repressed state to the fully active state): see Figure S.5. We again illustrate how to use the algorithm for the 4D model when $\mathrm{D_{tot}} = 2$; for larger $\mathrm{D_{tot}}$, the methodology will be the same. The explanation of the panels for Input, Step 1 and Step 2 is given above with $\mathcal{B} = \{(0,2,0,0)^T\}$. Step 3 involves two iterations. In the first iteration, the collection of the nodes $(1,0,1,0)^T$, $(1,0,0,1)^T$ and $(2,0,0,0)^T$ (called an r-connected set C_1) is condensed into a single node c_1. The order of the pole of the sojourn time in C_1 is $p(c_1) = \max_{u\in C_1} p(u) + \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u,v) \in E\} = 1 + 1 = 2$, where E denotes the edge set of the graph in Step 3 before the first iteration. Moreover, the value of $\mathcal{K}$ of edges between c_1 and an original node w that is not in C_1 are defined if there is an edge between some node $u \in C_1$ and w and according to the formula $\mathcal{K}_{c_1,w} = \min\{\mathcal{K}_{uw} : u \in C_1 \text{ and } (u,w) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_1, v \notin C_1 \text{ and } (u,v) \in E\}$, $\mathcal{K}_{w,c_1} = \min\{\mathcal{K}_{w,v} : v \in C_1 \text{ and } (w,v) \in E\}$. In the second iteration of Step 3, the collection of all nodes except for $(0,0,0,0)^T$ and a (called an r-connected set C_2) is condensed to a single node c_2. The order of the pole of the sojourn time in C_2 is $p(c_2) = \max_{u\in C_2} p(u) + \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 2 + 0 = 2$, where E denotes the edge set of the graph in Step 3 before the second iteration. Moreover, $\mathcal{K}_{c_2,(0,0,0,0)^T} = \min\{\mathcal{K}_{u,(0,0,0,0)^T} : u \in C \text{ and } (u,(0,0,0,0)^T) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 1 - 0 = 1$, $\mathcal{K}_{c_2,a} = \min\{\mathcal{K}_{ua} : u \in C_2 \text{ and } (u,a) \in E\} - \min\{\mathcal{K}_{uv} : u \in C_2, v \notin C_2 \text{ and } (u,v) \in E\} = 0 - 0 = 0$, and $\mathcal{K}_{(0,0,0,0)^T,c_2} = \min\{\mathcal{K}_{(0,0,0,0)^T,v} : v \in C_2 \text{ and } ((0,0,0,0)^T, v) \in E\} = 0$. Step 4 involves two iterations. In the first iteration, we fix the node with the largest value of p, which is c_2 in our case. At any node other than a that is connected to c_2 (i.e., the origin $(0,0,0,0)^T$), the value of p is updated to $p((0,0,0,0)^T) = \max\{p((0,0,0,0)^T), p(c_2) - \mathcal{K}_{(0,0,0,0)^T,c_2}\} = \max\{0, 2-0\} = 2$, and then any edges leading to or from c_2 are removed. In the second iteration, of the remaining nodes, we fix the node with the largest value of p, which is the origin. When all of the nodes other than a have been fixed, the order of the pole of the mean first passage time from each state in $\mathcal{B}^c$ to $\mathcal{B}$ is given by the fixed value of the node to which the state belongs.

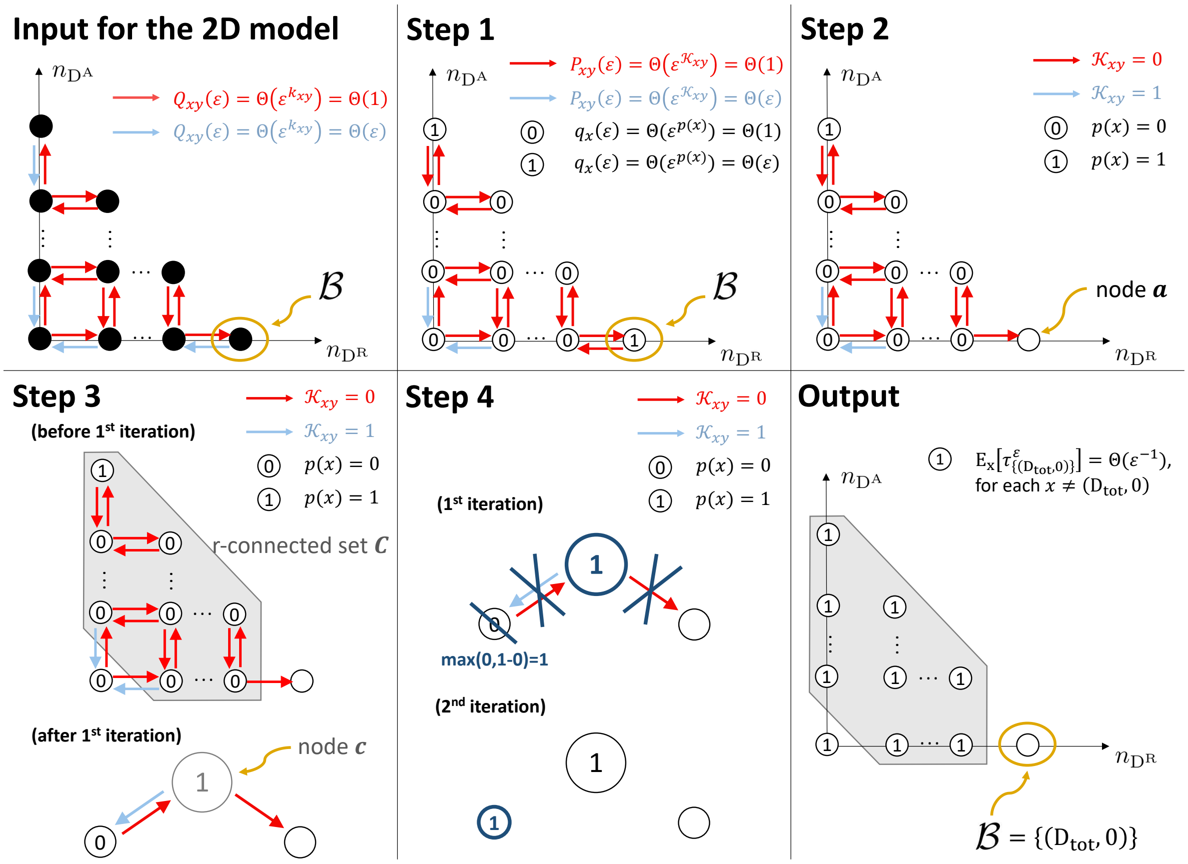

Figure S.1: **Key steps of the algorithm for the 2D model.**

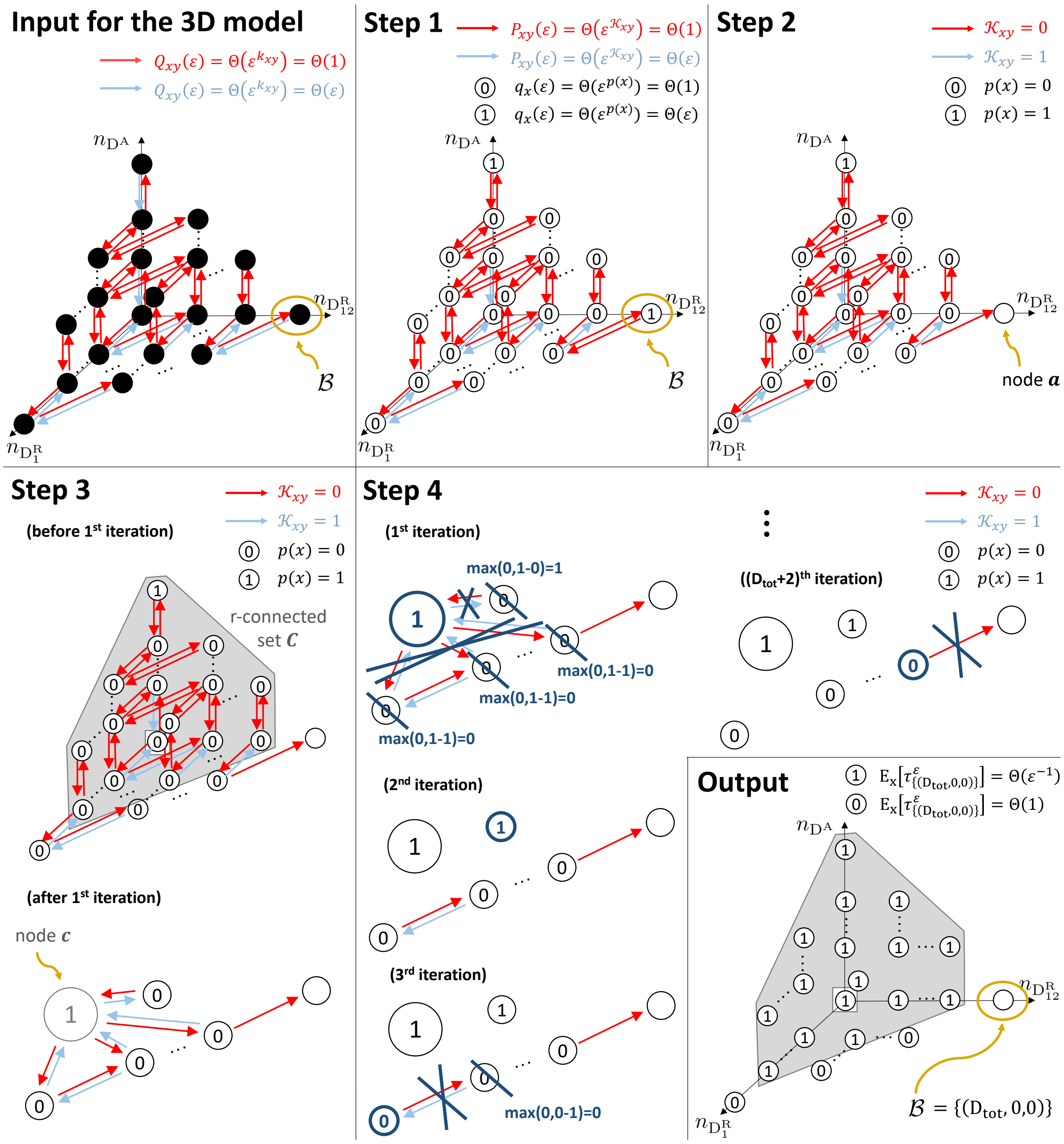

Figure S.2: **Key steps of the algorithm for the 3D model (from the fully active state to the fully repressed state).**

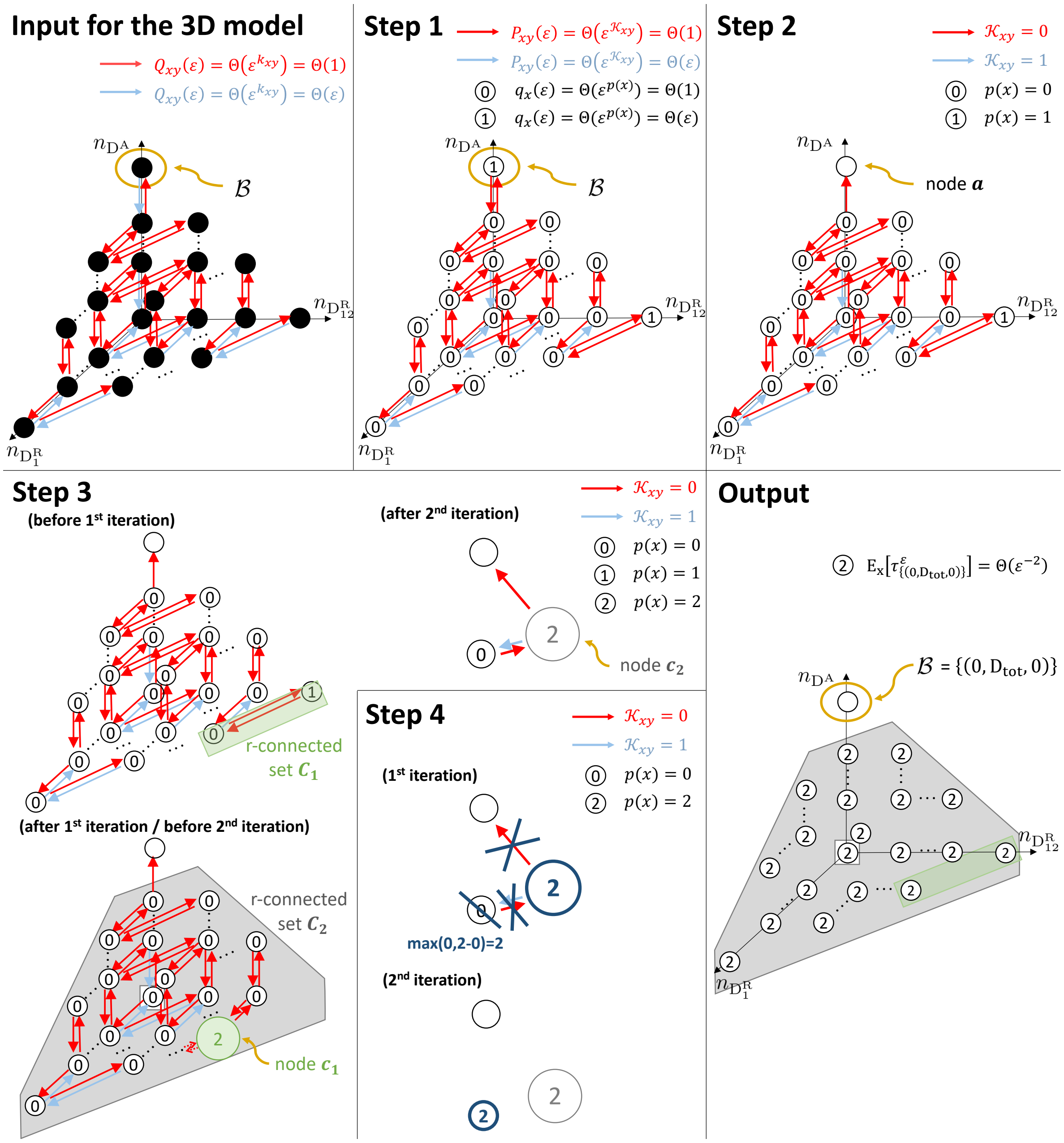

Figure S.3: **Key steps of the algorithm for the 3D model (from the fully repressed state to the fully active state).**

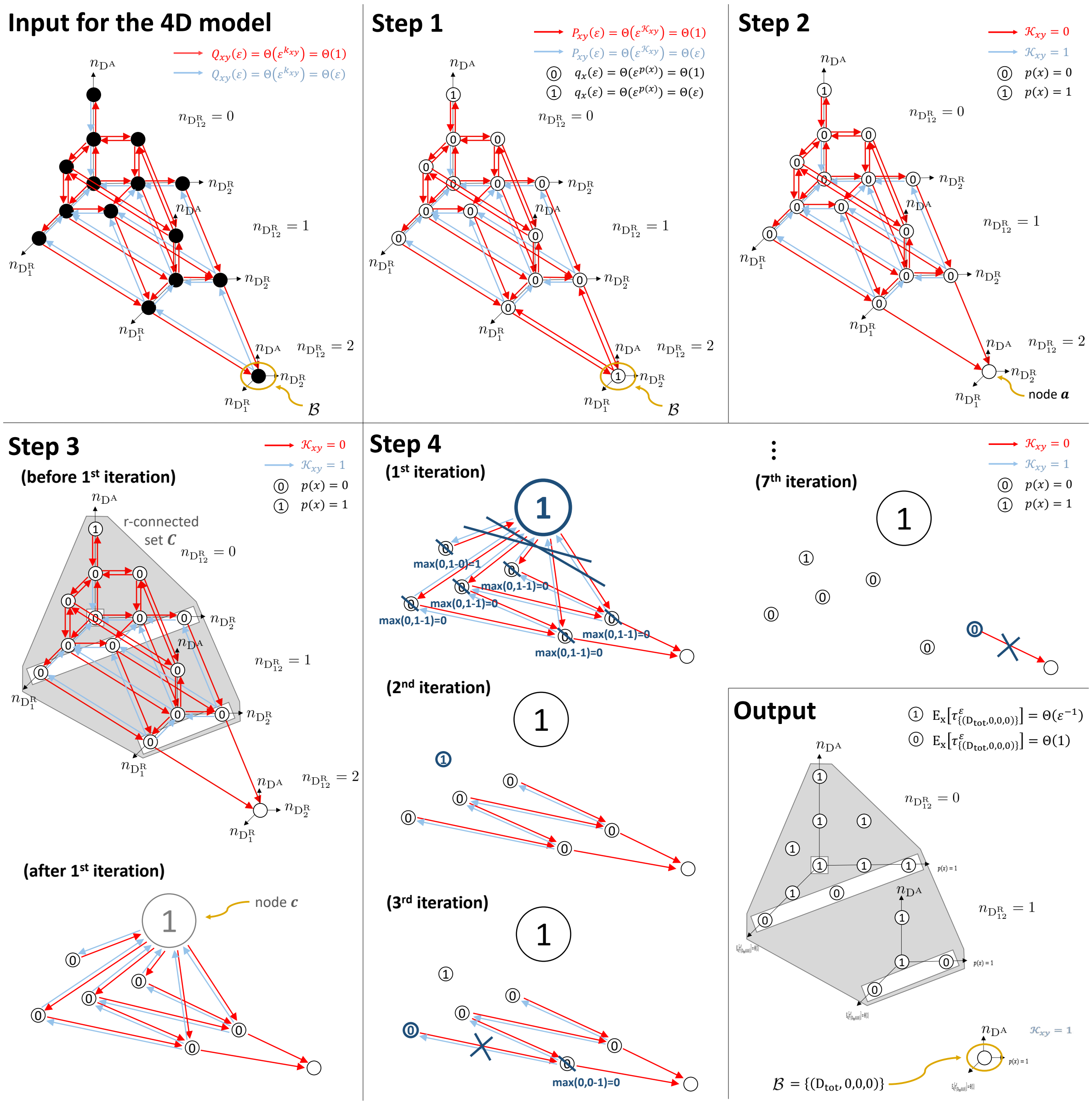

Figure S.4: **Key steps of the algorithm for the 4D model (from the fully active state to the fully repressed state).**

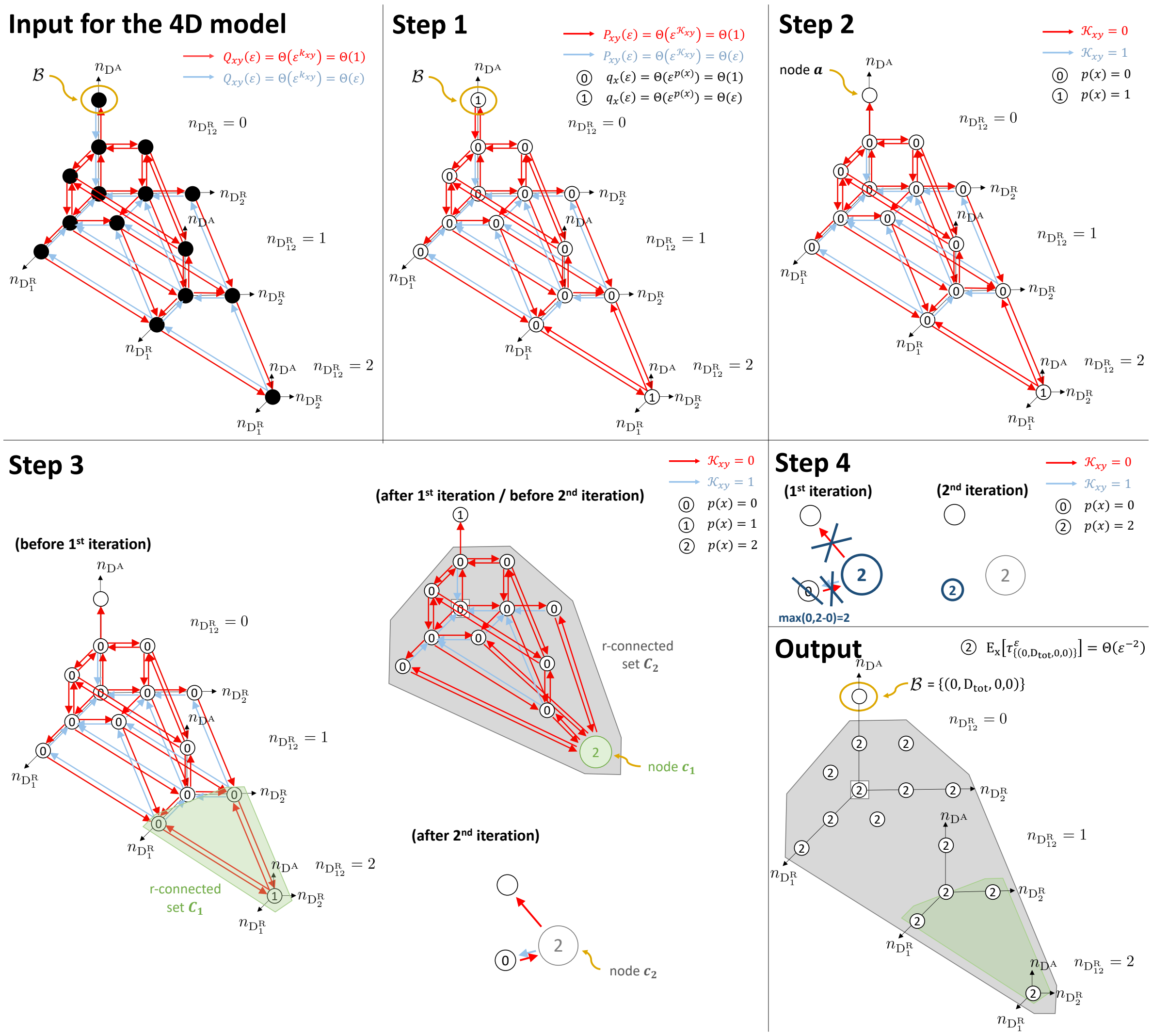

Figure S.5: **Key steps of the algorithm for the 4D model (from the fully repressed state to the fully active state).**

S.7 Leading coefficient for the MFPT

S.7.1 Proof of Theorem 4.7

Proof. Fix $\lambda = \max\{q_x(\varepsilon) : x \in \mathcal{X}, 0 \leq \varepsilon < \varepsilon_0\}$. The λ here should not be confused with other rates λ with subscripts and/or superscripts used elsewhere of this paper. In the following, we use the breve symbol to denote notation associated with discrete time Markov chains defined below.

Let $Y^\varepsilon = \{Y^\varepsilon(n) : n \in \mathbb{Z}_+\}$ be a discrete time Markov chain with transition matrix $\breve{P}(\varepsilon) = I + \frac{1}{\lambda}Q(\varepsilon)$ for each $0 \leq \varepsilon < \varepsilon_0$[5]. Note that Y^ε is a singularly perturbed discrete time Markov chain under the definition of Avrachenkov et al. [2]. Let $\breve{\Pi}(\varepsilon)$ be the ergodic projection of Y^ε and $\breve{H}(\varepsilon)$ be the deviation matrix of Y^ε (see definitions in SI - Section S.7.2). The ergodic projection of Y^0 is $\breve{\Pi}(0) = \breve{W}\breve{M}$, where I is the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix and

$$\breve{W} = \left(\frac{I}{-T_0^{-1}R_0}\right) \quad \text{and} \quad \breve{M} = \left(\, I \mid 0 \, \right).$$

Then,

$$\breve{M}\left(\frac{1}{\lambda}Q^{(1)}\right)\breve{W} = \frac{1}{\lambda}\left(\, I \mid 0 \, \right)\left(\begin{array}{c|c} A_1 & S_1 \\ R_1 & T_1 \end{array}\right)\left(\frac{I}{-T_0^{-1}R_0}\right) = \frac{1}{\lambda}(A_1 + S_1(-T_0)^{-1}R_0) = \frac{1}{\lambda}Q_{\mathcal{A}}.$$

Assumptions 4.1, 4.2 and Lemma S.1 imply that the null space of this matrix is one dimensional.

Using the computational algorithm in Section 6.3.1 of [2], the generator[6] for an aggregated discrete time Markov chain is $\breve{M}\left(\frac{1}{\lambda}Q^{(1)}\right)\breve{W} = \frac{1}{\lambda}Q_{\mathcal{A}}$, whose null space is one dimensional. Then, by the computational algorithm on page 176-177 of [2] the deviation matrix $\breve{H}(\varepsilon)$ has a Laurent series expansion with order of the pole equal to one:

$$\breve{H}(\varepsilon) = \frac{1}{\varepsilon}\breve{H}^{(-1)} + \breve{H}^{(0)} + \varepsilon\breve{H}^{(1)} + \ldots.$$

Since the aggregated Markov chain has a single recurrent class by Assumption 4.2, the ergodic projection of the aggregated Markov chain is $\mathbb{1}\alpha$, where α is a row vector denoting the unique stationary distribution of the aggregated discrete time Markov chain. The deviation matrix of this aggregated Markov chain is $\breve{D} = (-\frac{1}{\lambda}Q_{\mathcal{A}} + \mathbb{1}\alpha)^{-1} - \mathbb{1}\alpha$. By Theorem 6.7 in [2],

$$\breve{H}^{(-1)} = \breve{W}\breve{D}\breve{M} = \left(\frac{I}{-T_0^{-1}R_0}\right)\breve{D}\left(\, I \mid 0 \, \right) = \left(\begin{array}{c|c} \breve{D} & 0 \\ \hline T_0^{-1}R_0\breve{D} & 0 \end{array}\right).$$

For each $0 < \varepsilon < \varepsilon_0$, let $\breve{h}_{x,y}(\varepsilon)$ be the mean first passage time from x to y in Y^ε. Then, the

[5] In general, when $0 < \varepsilon < \varepsilon_0$, the discrete time Markov chain Y^ε is different from the embedded discrete time Markov chain described in Section 3.1. In particular, the discrete time Markov chain used here can have self loops, whereas the embedded discrete time Markov chain has no self loops.

[6] The transition matrix for a discrete time Markov chain with generator $\mathcal{G}$ is $\mathcal{P} = I + \mathcal{G}$.

mean first passage time from x to y in X^ε is

$$\begin{aligned} h_{x,y}(\varepsilon) = \frac{1}{\lambda}\breve{h}_{x,y}(\varepsilon) &= \frac{1}{\lambda}\frac{\breve{H}_{y,y}(\varepsilon) - \breve{H}_{x,y}(\varepsilon)}{\pi_y(\varepsilon)} \\ &= \frac{1}{\lambda}\frac{(\frac{1}{\varepsilon}\breve{D}_{y,y} + O(1)) - (\frac{1}{\varepsilon}\breve{D}_{x,y} + O(1))}{\pi_y^{(k_y)}\varepsilon_y^{k_y} + O(\varepsilon^{k_y+1})} \\ &= \frac{D_{y,y} - D_{x,y}}{\pi_y^{(k_y)}}\frac{1}{\varepsilon^{k_y+1}} + O\left(\frac{1}{\varepsilon^{k_y}}\right), \end{aligned}$$

where we used (S.56) to show that $\breve{D} = \lambda D$. The above equations use the properties of the deviation matrix given in SI - Section S.7.2.

When $\hat{X}_{\mathcal{A}}$ is irreducible, then $\pi^{(0)} = \alpha$ has all strictly positive entries. Again, by SI - Section S.7.2, the mean first passage time from x to y in $\hat{X}_{\mathcal{A}}$ is finite and positive, and it is $\frac{1}{\lambda}\frac{\breve{D}_{y,y}-\breve{D}_{x,y}}{\alpha_y} = \frac{D_{y,y}-D_{x,y}}{\pi_y^{(0)}}$. In this case, the order of the pole of $h_{x,y}(\varepsilon)$ is one and the leading coefficient is the mean first passage time from x to y in $\hat{X}_{\mathcal{A}}$. ■

S.7.2 Properties of the deviation matrix for a discrete time Markov chain

In this section, we will start with a few results stated in Section 6.1 of Avrachenkov et al. [2] about discrete time Markov chains with finite state space. These include the definitions and properties of the ergodic projection, the fundamental matrix and the deviation matrix. Then, we show one more fact about the deviation matrix. Lastly, Theorem 4.4.7 of Kemeny and Snell [14] gave a formula for mean first passage times for irreducible discrete time Markov chains in terms of the fundamental matrix and the stationary distribution, which is also briefly mentioned in [2]. We will write this in terms of the deviation matrix and the stationary distribution with a simple modification.

Suppose $Y = \{Y(n) : n \in \mathbb{Z}_+\}$ is a discrete time Markov chain with a finite state space $\mathcal{Y}$. Suppose the state space $\mathcal{Y}$ is partitioned into m ergodic classes (possibly including absorbing states) and a set of transient states, and accordingly, the transition matrix $\breve{P}$ is

$$\breve{P} = \left(\begin{array}{ccc|c} \breve{A}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \breve{A}_m & 0 \\ \hline \breve{R}_1 & \dots & \breve{R}_m & \breve{T} \end{array}\right).$$

The ergodic projection of Y is given by the *Cesaro* limit,

$$\breve{\Pi} = \lim_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} \breve{P}^n.$$

It follows that $\breve{\Pi}(I - \breve{P}) = 0$ and $\breve{\Pi}\breve{\Pi} = \breve{\Pi}$. The ergodic projection $\breve{\Pi}$ is the eigenprojection of the transition matrix $\breve{P}$ corresponding to its maximal eigenvalue 1. That is, if $\breve{\pi}_i$ is the unique stationary distribution for the discrete time Markov chain with transition matrix $\breve{A}_i$

for $1 \le i \le m$, then $\breve{\Pi} = \breve{W}\breve{M}$ with

$$\breve{W} = \left(\begin{array}{ccc} \mathbb{1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbb{1} \\ \hline (I-\breve{T})^{-1}\breve{R}_1\mathbb{1} & \dots & (I-\breve{T})^{-1}\breve{R}_m\mathbb{1} \end{array} \right) \quad \text{and} \quad \breve{M} = \left(\begin{array}{ccc|c} \breve{\pi}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \breve{\pi}_m & 0 \end{array} \right),$$

where $\breve{W}$ and $\breve{M}$ form bases for the right and left eigenspaces, respectively, which implies that $\breve{P}\breve{W} = \breve{W}$ and $\breve{M}\breve{P} = \breve{M}$. One can see that $v(I - \breve{P} + \breve{\Pi}) = 0$ implies that $v = 0$ and so $(I - \breve{P} + \breve{\Pi})$ is invertible. The fundamental matrix $\breve{Z}$ and the deviation matrix $\breve{H}$ of Y are well-defined:

$$\breve{Z} = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \sum_{\ell=0}^{n} (\breve{P} - \breve{\Pi})^{\ell} = (I - \breve{P} + \breve{\Pi})^{-1},$$

$$\breve{H} = \breve{Z} - \breve{\Pi} = (I - \breve{P} + \breve{\Pi})^{-1} - \breve{\Pi}.$$

We also have that $\breve{H}\breve{\Pi} = (\breve{Z} - \breve{\Pi})\breve{\Pi} = 0$ since $\breve{P}\breve{\Pi} = \breve{P}\breve{W}\breve{M} = \breve{W}\breve{M} = \breve{\Pi}$, $\breve{\Pi} = \breve{Z}(I - \breve{P} + \breve{\Pi})\breve{\Pi} = \breve{Z}(\breve{\Pi} - \breve{\Pi} + \breve{\Pi}) = \breve{Z}\breve{\Pi}$, and $\Pi^2 = \Pi$.

Now, we show a property of the deviation matrix that is not in [2] and is useful in Section 4.2.2. Suppose Q is an infinitesimal generator for a continuous time Markov chain on $\mathcal{Y}$ and $|Q_{y,y}| \le \lambda$ for all $y \in \mathcal{Y}$. Then, $\breve{P} = I + \frac{1}{\lambda}Q$ defines a transition matrix for a discrete time Markov chain. The associated ergodic projection and deviation matrix $\breve{\Pi}$ and $\breve{H}$ for $\breve{P}$ satisfy $\breve{\Pi}Q = \lambda(\breve{\Pi}\breve{P} - \breve{\Pi}) = 0$ and

$$\breve{H} = \left(I - \left(I + \frac{1}{\lambda}Q\right) + \breve{\Pi}\right)^{-1} - \breve{\Pi} = \left(-\frac{1}{\lambda}Q + \breve{\Pi}\right)^{-1} - \breve{\Pi},$$

and so

$$\begin{aligned} I &= (\breve{H} + \breve{\Pi})\left(-\frac{1}{\lambda}Q + \breve{\Pi}\right) = -\frac{1}{\lambda}\breve{H}Q + \breve{H}\breve{\Pi} - \frac{1}{\lambda}\breve{\Pi}Q + \breve{\Pi}\breve{\Pi} \\ &= -\frac{1}{\lambda}\breve{H}Q + \breve{\Pi} = -\frac{1}{\lambda}\breve{H}Q + \frac{1}{\lambda}\breve{H}\breve{\Pi} - \breve{\Pi}Q + \breve{\Pi}\breve{\Pi} = \left(\frac{1}{\lambda}\breve{H} + \breve{\Pi}\right)(-Q + \breve{\Pi}). \end{aligned}$$

Thus,

$$\frac{1}{\lambda}\breve{H} = ((-Q + \breve{\Pi})^{-1} - \breve{\Pi}), \tag{S.56}$$

where we have used the fact that $-Q + \breve{\Pi}$ is invertible because $\breve{\Pi}$ is the eigenprojection of Q corresponding to the eigenvalue 0.

Lastly, assume that Y is irreducible. Then, Y has a unique stationary distribution $\breve{\pi}$, which is a row vector, and the ergodic projection of Y is $\mathbb{1}\breve{\pi}$. By Theorem 4.4.7 in [14], the mean first passage time from $x \in \mathcal{Y}$ to $y \in \mathcal{Y}$ is $\frac{\breve{Z}_{y,y} - \breve{Z}_{x,y}}{\breve{\pi}_y}$. Since $\breve{Z} = \breve{H} + \mathbb{1}\breve{\pi}$, the mean first passage time from $x \in \mathcal{Y}$ to $y \in \mathcal{Y}$ is

$$\frac{(\breve{H}_{y,y} + (\mathbb{1}\breve{\pi})_{y,y}) - (\breve{H}_{x,y} + (\mathbb{1}\breve{\pi})_{x,y})}{\breve{\pi}_y} = \frac{(\breve{H}_{y,y} + \breve{\pi}_y) - (\breve{H}_{x,y} + \breve{\pi}_y)}{\breve{\pi}_y} = \frac{\breve{H}_{y,y} - \breve{H}_{x,y}}{\breve{\pi}_y}.$$

S.8 1D Model: additional mathematical details

Verification of Assumption 4.1. In order to show that Assumption 4.1 holds, consider the states $a = 0$ and $r = \mathrm{D_{tot}}$ and the set $\mathcal{T} = \{1, \ldots, \mathrm{D_{tot}} - 1\}$ defined in Section 2.1. Since $\mathrm{D_{tot}} \geq 2$, $\mathcal{T} \neq \emptyset$. From (2.2), we can see that $Q_{a,a+1}(0) = Q_{a,a-1}(0) = Q_{r,r+1}(0) = Q_{r,r-1}(0) = 0$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x \in \mathcal{T}$. Since $Q_{z,z+1}(0) = \frac{k_E^A}{V}(\mathrm{D_{tot}} - z)z > 0$ for all $z \in \{1, \ldots, \mathrm{D_{tot}} - 1\}$, we have $Q_{x,x+1}(0) \ldots Q_{\mathrm{D_{tot}}-1,\mathrm{D_{tot}}}(0) > 0$. By Lemma S.8 and the fact that r is an absorbing state, we have that x is a transient state for X^0.

Verification of Assumptions 4.4 and 4.2. By Lemma 4.4, it suffices to show Assumption 4.4 holds. From (2.2), we can see that $\tilde{Q}_{a,a+1} > 0$. From the analysis made to prove Assumption 4.1, we know that there is a positive probability for $\tilde{X}$ to transition from $x \in \mathcal{X} \setminus \{a, r\}$ to r. It follows that any state $x \in \mathcal{X} \setminus \{r\}$ leads to r under $\tilde{X}$. Now, we would like to show that there is a positive probability for transition from r to $x \neq r \in \mathcal{X}$ for the process $\tilde{X}$. This is because $\tilde{Q}_{r,r-1} = b\frac{k_E^A}{V}\mathrm{D}_{\mathrm{tot}}^2 > 0$ and $\tilde{Q}_{z,z-1} = Q_{z,z-1}(0) = \mu\frac{k_E^A}{V}(\mathrm{D_{tot}} - z)z > 0$ for all $z \in \{1, \ldots, \mathrm{D_{tot}} - 1\}$. Thus, r leads to any state in $\mathcal{X} \setminus \{r\}$ under $\tilde{X}$. Combining the above, we see that $\tilde{X}$ is irreducible and Assumption 4.4 holds.

Stationary distribution. Let us consider a one-dimensional finite state continuous time Markov chain in which the state space $\mathcal{X} = \{0, 1, \ldots, K\}$ and the off-diagonal entries of the infinitesimal generator Q are all zero except for the following positive rates:

$$\text{(S.57)} \qquad \begin{aligned} Q_{x,x+1} &= \lambda_x \quad \text{if } x \in \{0, \ldots, K-1\}, \\ Q_{x,x-1} &= \gamma_x \quad \text{if } x \in \{1, \ldots, K\}. \end{aligned}$$

Thus, the continuous time Markov chain is a birth-and-death process, it satisfies detailed balance (see [7]) and so the stationary distribution $\pi = (\pi_x)_{x \in \{0,1,\ldots,K\}}$ satisfies

$$\pi_x = \frac{\lambda_{x-1}}{\gamma_x}\pi_{x-1}, \qquad \text{for } x \in \{1, \ldots, K\}.$$

Applying this equality recursively, we can express π_x, $x \in \{1, \ldots, K\}$, as a function of π_0, obtaining

$$\text{(S.58)} \qquad \pi_x = \pi_0 \prod_{i=1}^{x} \frac{\lambda_{i-1}}{\gamma_i}.$$

Using the fact that $\sum_{j=0}^{K} \pi_j = 1$, we obtain

$$\text{(S.59)} \qquad \pi_0 = \frac{1}{1 + \sum_{j=1}^{K} \left(\prod_{i=1}^{j} \frac{\lambda_{i-1}}{\gamma_i} \right)}.$$

Substituting (S.59) in (S.58), we obtain

$$\pi_x = \frac{\prod_{i=1}^{x} \frac{\lambda_{i-1}}{\gamma_i}}{1+\sum_{j=1}^{K}\left(\prod_{i=1}^{j}\frac{\lambda_{i-1}}{\gamma_i}\right)} \qquad \text{for } x \in \{1,\dots,K\}. \tag{S.60}$$

Now, consider the one-dimensional continuous time Markov chain introduced in Section 2.1 with state space $\mathcal{X} = \{0, 1, \dots, \mathrm{D_{tot}}\}$ and infinitesimal generator as defined in (2.2), which has nonzero off-diagonal entries given, for $\varepsilon > 0$, by

$$\begin{aligned} \lambda_x^{\varepsilon} &:= Q_{x,x+1}(\varepsilon) = \left(\frac{k_E^A}{V}x + \varepsilon\frac{k_E^A}{V}\mathrm{D_{tot}}\right)(\mathrm{D_{tot}}-x) \qquad \text{if } x \in \{0,\dots,\mathrm{D_{tot}}-1\},\\ \gamma_x^{\varepsilon} &:= Q_{x,x-1}(\varepsilon) = \mu\left(\frac{k_E^A}{V}(\mathrm{D_{tot}}-x) + b\varepsilon\frac{k_E^A}{V}\mathrm{D_{tot}}\right)x \qquad \text{if } x \in \{1,\dots,\mathrm{D_{tot}}\}. \end{aligned} \tag{S.61}$$

By substituting the expressions for the rates in (S.61) into (S.59)-(S.60), and suitably rearranging the terms, we obtain that

$$\pi_x(0) = \lim_{\varepsilon\to 0}\pi_x(\varepsilon) = \begin{cases} \frac{b\mu^{\mathrm{D_{tot}}}}{1+b\mu^{\mathrm{D_{tot}}}} & \text{if } x = 0\\ 0 & \text{if } x \in \{1,\dots,\mathrm{D_{tot}}-1\}\\ \frac{1}{1+b\mu^{\mathrm{D_{tot}}}} & \text{if } x = \mathrm{D_{tot}}. \end{cases}$$

Mean first passage time. Consider the one-dimensional, finite state, continuous time Markov chain introduced in (S.57). We will determine an analytical expression for the MFPT from $x = K$ to $x = 0$ and from $x = 0$ to $x = K$ for this chain. We first focus on the former. For this, we exploit first step analysis (see Equation 3.1 of [16]), proceeding in a similar manner to that for (3.2), to obtain

$$\begin{cases} h_{0,0} = 0,\\ h_{x,0} = \frac{1}{\lambda_x+\gamma_x} + \frac{\lambda_x}{\lambda_x+\gamma_x}h_{x+1,0} + \frac{\gamma_x}{\lambda_x+\gamma_x}h_{x-1,0} & \text{if } x \in \{1,\dots,K-1\},\\ h_{K,0} = \frac{1}{\gamma_K} + h_{K-1,0}, \end{cases} \tag{S.62}$$

where for $x, y \in \mathcal{X}$, $h_{x,y} = \mathbb{E}_x[\tau_y]$, $\tau_y = \inf\{t \geq 0 : X(t) = y\}$, X is the continuous time Markov chain with infinitesimal generator given by (S.57). Now, defining $\Delta h_{x,x-1} = h_{x,0} - h_{x-1,0}$ for $x \in \{1, ..., K\}$, we can rewrite (S.62) in the following way:

$$\begin{cases} h_{0,0} = 0,\\ \Delta h_{x,x-1} = \frac{1}{\gamma_x} + \frac{\lambda_x}{\gamma_x}\Delta h_{x+1,x} & \text{if } x \in \{1,\dots,K-1\},\\ \Delta h_{K,K-1} = \frac{1}{\gamma_K}. \end{cases} \tag{S.63}$$

From (S.63), we have an explicit formula for $\Delta h_{K,K-1}$ and any $\Delta h_{x,x-1}$ can be expressed as a function of $\Delta h_{x+1,x}$. Furthermore, if we sum the $\Delta h_{x,x-1}$ for $x = 1, \dots, K$, we obtain

$$h_{K,0} = h_{K,0} - h_{0,0} = \sum_{x=1}^{K}(\Delta h_{x,x-1}) = \Delta h_{1,0} + \Delta h_{2,1} + ... + \Delta h_{K-1,K-2} + \Delta h_{K,K-1}. \tag{S.64}$$

Thus, to evaluate the MFPT from $x = K$ to $x = 0$, we can calculate $\Delta h_{x,x-1}$ for $x = K, K-1, ..., 1$ and then sum all of the terms. Defining $r_j = \frac{\lambda_1\lambda_2...\lambda_j}{\gamma_1\gamma_2...\gamma_j}$, for $j = 1, ..., K$, we obtain

$$
\begin{aligned}
h_{K,0} &= \frac{1}{\gamma_K}\left(1 + \frac{\lambda_{K-1}}{\gamma_{K-1}} + \frac{\lambda_{K-1}\lambda_{K-2}}{\gamma_{K-1}\gamma_{K-2}} + ... + r_{K-1}\right) \\
&\quad + \frac{1}{\gamma_{K-1}}\left(1 + \frac{\lambda_{K-2}}{\gamma_{K-2}} + \frac{\lambda_{K-2}\lambda_{K-3}}{\gamma_{K-2}\gamma_{K-3}} + ... + r_{K-2}\right) + ... + \frac{1}{\gamma_1} \\
&= \frac{r_{K-1}}{\gamma_K}\left(1 + \sum_{i=1}^{K-1}\frac{1}{r_i}\right) + \sum_{i=2}^{K-1}\left[\frac{r_{i-1}}{\gamma_i}\left(1 + \sum_{j=1}^{i-1}\frac{1}{r_j}\right)\right] + \frac{1}{\gamma_1}.
\end{aligned} \tag{S.65}
$$

With a similar procedure, we can obtain the MFPT from $x = 0$ to $x = K$. More precisely, defining $\tilde{r}_j = \frac{\gamma_{K-1}\gamma_{K-2}...\gamma_{K-j}}{\lambda_{K-1}\lambda_{K-2}...\lambda_{K-j}}$, we have

$$
\begin{aligned}
h_{0,K} &= \frac{1}{\lambda_0}\left(1 + \frac{\gamma_1}{\lambda_1} + \frac{\gamma_1\gamma_2}{\lambda_1\lambda_2} + ... + \tilde{r}_{K-1}\right) \\
&\quad + \frac{1}{\lambda_1}\left(1 + \frac{\gamma_2}{\lambda_2} + \frac{\gamma_2\gamma_3}{\lambda_2\lambda_3} + ... + \tilde{r}_{K-2}\right) + ... + \frac{1}{\lambda_{K-1}} \\
&= \frac{\tilde{r}_{K-1}}{\lambda_0}\left(1 + \sum_{j=1}^{K-1}\frac{1}{\tilde{r}_i}\right) + \sum_{i=2}^{K-1}\left[\frac{\tilde{r}_{i-1}}{\lambda_{K-i}}\left(1 + \sum_{j=1}^{i-1}\frac{1}{\tilde{r}_j}\right)\right] + \frac{1}{\lambda_{K-1}}.
\end{aligned} \tag{S.66}
$$

A more detailed derivation of the $h_{0,K}$ and $h_{K,0}$ is given in [3].

Let us consider the one-dimensional continuous time Markov chain introduced in Section 2.1, with state space $\mathcal{X} = \{0, 1, \dots, \mathrm{D_{tot}}\}$ and infinitesimal transition rates that can be written as in (S.61). Since all of the transition rates are $O(1)$, except for λ_0^{ε} and $\gamma_{\mathrm{D_{tot}}}^{\varepsilon}$ which are $O(\varepsilon)$, then both $h_{\mathrm{D_{tot}},0}(\varepsilon)$ and $h_{0,\mathrm{D_{tot}}}(\varepsilon)$ are $O(1/\varepsilon)$. This means that in the limit as $\varepsilon \to 0$, $h_{\mathrm{D_{tot}},0}(\varepsilon)$ and $h_{0,\mathrm{D_{tot}}}(\varepsilon)$, which correspond to the time to memory loss of the repressed and active states, respectively, tend to infinity. Substituting parameters in (S.65) and (S.66) yields (2.7) and (2.8), respectively.

S.9 2D Model: additional mathematical details

Verification of Assumption 4.1. In order to show that Assumption 4.1 holds, consider the states $a = (0, \mathrm{D_{tot}})^T$ and $r = (\mathrm{D_{tot}}, 0)^T$ and the set $\mathcal{T} = \{i_1, \dots, i_m\}$ defined in Section 4.1.2. From (2.10), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \le j \le 4$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2)^T \in \mathcal{T}$. First, suppose $x_1 \neq 0$. By having the one-step transition along $v_2 = (0, -1)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k_E^A}{V} z_2 x_1 > 0$ for all $z = (x_1, z_2)^T$ and $1 \le z_2 \le x_2$, and having one-step transition along $v_3 = (1, 0)^T$ occurring $\mathrm{D_{tot}} - x_1$ times where $Q_{z,z+v_3}(0) = (\mathrm{D_{tot}} - z_1)\left(k_{W0}^R + k_W^R + \frac{k_M^R}{V} z_1\right) > 0$ for all $z = (z_1, 0)^T$ and $x_1 \le z_1 \le \mathrm{D_{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By Lemma S.8 and the fact that r is an absorbing state, we have that x is a transient state for X^0. On the other hand, suppose $x_1 = 0$. Since $x = (0, x_2) \in \mathcal{T}$, we have $0 \leq x_2 \leq \mathrm{D_{tot}} - 1$. We can first have a one-step transition along $v_3 = (1,0)^T$, where $Q_{x,x+v_3}(0) = (\mathrm{D_{tot}} - x_2)\left(k^R_{W0} + k^R_W\right) > 0$, to reach the state $(1, x_2)^T$ and then take the steps for the $x_1 \neq 0$ case to reach r. In this way, there is a positive probability of transition from x to the absorbing state r under $Q(0)$, and thus x is transient by Lemma S.8.

Verification of Assumptions 4.3 and 4.2. By Lemma 4.4, it suffices to show Assumption 4.3 holds. From (2.10), we can see that $\tilde{Q}_{a,a+v_2} > 0$. From the analysis made to prove Assumption 4.1, we know that there is a positive probability to transition from all $x \in \mathcal{X}\backslash\{a, r\}$ to r. Now, we would like to show that there is a positive probability to transition from r to $x = (x_1, x_2)^T \in \mathcal{X} \setminus \{(0,0)^T\}$ for the process $\tilde{X}$. We first can have a one-step transition along $v_4 = (-1,0)^T$ where $Q^{(1)}_{r,r+v_4} = \mu b \frac{k^A_M}{V}\mathrm{D}^2_{\mathrm{tot}} > 0$, then have a one-step transition along $v_1 = (0,1)^T$ where $Q^{(0)}_{r+v_4,r+v_4+v_1} = k^A_{W0} + k^A_W > 0$, then have one-step transitions along $v_4 = (-1,0)^T$ occurring $\mathrm{D_{tot}} - x_1 - 1$ times where $Q^{(0)}_{z,z+v_4} = \mu\frac{k^A_E}{V} z_1 > 0$ for all $z = (z_1, 1)^T$ and $x_1 + 1 \leq z_1 \leq \mathrm{D_{tot}} - 1$. If $x_2 \neq 0$, we finally have one-step transitions along $v_1 = (0,1)^T$ occurring $x_2 - 1$ times where $Q^{(0)}_{z,z+v_1} = (\mathrm{D_{tot}} - (x_1 + z_2))\left(k^A_{W0} + k^A_W + \frac{k^A_M}{V} z_2\right) > 0$ for all $z = (x_1, z_2)^T$ and $1 \leq z_2 \leq x_2 - 1$; if $x_2 = 0$ and $x_1 \neq 0$, we will make a one-step transition along $v_2 = (0,-1)^T$ to $(x_1, 0)^T$ where $Q^{(0)}_{z,z+v_2} = \frac{k^A_E}{V} x_1 > 0$ with $z = (x_1, 1)^T$ and $x_1 \geq 1$. Therefore, we have that there is a positive probability of transition from r to each $x \in \mathcal{X} \setminus \{(0,0)^T\}$. Since $Q^{(0)}_{-v_j,(0,0)^T} = 0$ for $j \in \{1,2,3,4\}$ such that $-v_j \in \mathcal{X}$, we conclude that $\mathcal{C} = \mathcal{X} \setminus \{(0,0)^T\}$ is a closed communicating class under $\tilde{Q}$ and since it contains $\mathcal{A}$, Assumption 4.3 holds. Note that Assumption 4.4 does not hold.

Stationary distribution. Here, we derive the expression for $\pi^{(1)}_x$, $x \in \mathcal{T} = \{i_1, \ldots, i_m\}$, for the case $\mathrm{D_{tot}} = 2$. In this case $T(\varepsilon) = T_0 + \varepsilon T_1$, with

$$T_0 = \begin{pmatrix} -q_3 & k^R_{W0} + k^R_W & 0 & 0 \\ \mu\frac{k^A_E}{V} & -\frac{k^A_E}{V}(1+\mu) & 0 & \frac{k^A_E}{V} \\ 2(k^A_{W0} + k^A_W) & 0 & -q_5 & 2(k^R_{W0} + k^R_W) \\ 0 & k^A_{W0} + k^A_W & 0 & -q_6 \end{pmatrix},$$

$$T_1 = \begin{pmatrix} -2\frac{k^A_M}{V} & 0 & 2\frac{k^A_M}{V} & 0 \\ 2\frac{k^A_M}{V}\mu b & -2\frac{k^A_M}{V}(1+\mu b) & 0 & 2\frac{k^A_M}{V} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2\frac{k^A_M}{V}\mu b & -2\frac{k^A_M}{V} \end{pmatrix},$$

in which $q_3 = (k^A_{W0} + k^A_W + \frac{k^A_M}{V} + k^R_{W0} + k^R_W)$, $q_5 = 2(k^A_{W0} + k^A_W + k^R_{W0} + k^R_W)$ and $q_6 = (k^A_{W0} + k^A_W + k^R_{W0} + k^R_W + \frac{k^R_M}{V})$. Then, by (4.5), $\beta^{(1)} = \pi^{(1)}_x$, $x \in \mathcal{T}$, is given by $\beta^{(1)} = \alpha S_1(-T_0)^{-1}$, where $\alpha = (\pi^{(0)}_{(0,2)}, \pi^{(0)}_{(2,0)})$ was derived in Section 4.1.2 (Eq. (4.14)). After some calculations,

$\pi_x^{(1)}$ can be written for $\mathrm{D_{tot}} = 2$ as

$$\pi_{i_1}^{(1)} = \frac{4b\varepsilon^2\left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^A \mu^2 \left(\bar{k}_W^R(\bar{k}_W^R + \frac{k_M^R}{V}) + (\bar{k}_W^A + \frac{k_M^A}{V})\left((1+\mu)(\bar{k}_W^R + \frac{k_M^R}{V}) + \mu\frac{k_M^A}{V}\right)\right)}{d_1 d_2},$$

$$\pi_{i_2}^{(1)} = \frac{4b\varepsilon^2\left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^A \bar{k}_W^R \mu \left((\bar{k}_W^R + \frac{k_M^R}{V})(\bar{k}_W^R + \bar{k}_W^A + k_M^A) + \mu(\bar{k}_W^A + \frac{k_M^A}{V})(\bar{k}_W^R + \bar{k}_W^A + k_M^R)\right)}{\frac{k_E^A}{V} d_1 d_2},$$

$$\pi_{i_3}^{(1)} = 0,$$

$$\pi_{i_4}^{(1)} = \frac{4b\varepsilon^2\left(\frac{k_M^A}{V}\right)^2 \bar{k}_W^R \mu \left((\bar{k}_W^R + \frac{k_M^R}{V})\left((1+\mu)(\bar{k}_W^A + \frac{k_M^A}{V}) + \mu\bar{k}_W^R\right) + \mu\bar{k}_W^A(\bar{k}_W^A + \frac{k_M^A}{V})\right)}{d_1 d_2},$$

in which $\bar{k}_W^A = k_{W0}^A + k_W^A$, $\bar{k}_W^R = k_{W0}^R + k_W^R$, $d_1 = \varepsilon\frac{k_M^A}{V}(\bar{k}_W^R(\bar{k}_W^R + \frac{k_M^R}{V}) + b\bar{k}_W^A\mu^2(\bar{k}_W^A + \frac{k_M^A}{V}))$, and $d_2 = ((\bar{k}_W^A + \frac{k_M^A}{V})((1+\mu)(\bar{k}_W^R + \frac{k_M^R}{V}) + \mu\bar{k}_W^A) + \bar{k}_W^R(\bar{k}_W^R + \frac{k_M^R}{V}))$ and in which $i_1 = (0,1)^T, i_2 = (1,1)^T, i_3 = (0,0)^T$, and $i_4 = (0,2)^T$.

S.10 3D Model: additional mathematical details

Verification of Assumption 4.1. In order to show that Assumption 4.1 holds, consider the states $a = (0, \mathrm{D_{tot}}, 0)^T$ and $r = (\mathrm{D_{tot}}, 0, 0)^T$ and the set $\mathcal{T} = \{i_1, \ldots, i_m\}$ defined in Section 5.1.1. From (5.2), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \le j \le 6$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2, x_3)^T \in \mathcal{T}$. First, suppose $x_1 + x_3 \neq 0$. By having the one-step transitions along $v_2 = (0,-1,0)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k_E^A}{V}(x_3 + 2x_1)z_2 > 0$ for all $z = (x_1, z_2, x_3)^T$ and $1 \le z_2 \le x_2$, then having one-step transitions along $v_3 = (0,0,1)^T$ occurring $\mathrm{D_{tot}} - x_1 - x_3$ times where $Q_{z,z+v_3}(0) = (\mathrm{D_{tot}} - (x_1 + z_3))\left(k_{W0}^1 + k_W^1 + \frac{k_M^{'}}{V}x_1\right) > 0$ for all $z = (x_1, 0, z_3)^T$ and $x_3 \le z_3 \le \mathrm{D_{tot}} - x_1 - 1$ and finally having one-step transitions along $v_5 = (1,0,-1)^T$ occurring $\mathrm{D_{tot}} - x_1$ times where $Q_{z,z+v_5}(0) = (\mathrm{D_{tot}} - z_1)\left(k_{W0}^2 + \frac{k_M}{V}z_1 + \frac{\bar{k}_M}{V}\frac{\mathrm{D_{tot}}+z_1-1}{2}\right) > 0$ for all $z = (z_1, 0, \mathrm{D_{tot}} - z_1)^T$ and $x_1 \le z_1 \le \mathrm{D_{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By Lemma S.8 and the fact that r is an absorbing state, we have that x is a transient state for X^0. On the other hand, suppose $x_1 + x_3 = 0$. Since $x = (0, x_2, 0)^T \in \mathcal{T}$, we have $0 \le x_2 \le \mathrm{D_{tot}} - 1$. We can first have a one-step transition along $v_3 = (0,0,1)^T$, where $Q_{x,x+v_3}(0) = (\mathrm{D_{tot}} - x_2)\left(k_{W0}^1 + k_W^1\right) > 0$, to reach the state $(0, x_2, 1)^T$ and then take the steps in the $x_1 + x_3 \neq 0$ case. In this way, there is a positive probability of transition from x to the absorbing state r, and thus x is transient by Lemma S.8.

Verification of Assumption 4.2. To show that Assumption 4.2 holds, consider the continuous time Markov chain $\tilde{X}$ with infinitesimal generator $\tilde{Q}$ as described in (4.7) and shown in Fig. 6(d). We will first see that $\{i_m, r\}$ forms a closed class under $\tilde{Q}$. For this, we see that $Q_{r,r+v_j}(\varepsilon)$ vanishes for every $1 \le j \le 5$ and $\varepsilon \ge 0$, while $Q_{r,r+v_6}(\varepsilon) = \varepsilon\mu b\frac{k_M^A}{V}\mathrm{D_{tot}^2}$. Therefore, the only transition from r under $\tilde{Q}$ is given by $\tilde{Q}_{r,r+v_6} = \mu b\frac{k_M^A}{V}\mathrm{D_{tot}^2} > 0$, where $r + v_6 = i_m$. From (5.2), we can see that $Q_{i_m,i_m+v_j}(0) = 0$ for every $j \in \{1,2,3,4,6\}$ and $Q_{i_m,i_m+v_5}(0) = k_{W0}^2 + \frac{k_M}{V}(\mathrm{D_{tot}} - 1) + \frac{\bar{k}_M}{V}(\mathrm{D_{tot}} - 1) > 0$. Since $i_m + v_5 = r$ we see that

$\tilde{Q}_{i_m,r} > 0$. Therefore, $\{i_m, r\}$ forms a closed class under $\tilde{Q}$. The fact that $\hat{X}_{\mathcal{A}}$, shown in Fig. 6(e) consists of erasing the times from $\tilde{X}$ in which the process is in $\mathcal{T}$, together with Lemma 4.3, yields that r is an absorbing state under $Q_{\mathcal{A}}$. From (5.2), we can see that $\tilde{Q}_{a,i_1} > 0$. From the analysis made to prove Assumption 4.1 we obtain that i_1 leads to r under $\tilde{Q}$ which is part of a closed class. By interpreting $\hat{X}_{\mathcal{A}}$ again as a time-change of $\tilde{X}$, by Lemma 4.3 we obtain that a is transient under $Q_{\mathcal{A}}$. As a consequence, $Q_{\mathcal{A}}$ has a single recurrent class consisting of the state r, and so Assumption 4.2 holds, and furthermore, $\alpha = [\alpha_a, \alpha_r]$ with $\alpha_a = 0$ and $\alpha_r = 1$. In addition, the previous arguments show that neither Assumption 4.3 nor 4.4 holds for this model.

Stationary distribution. Here, we derive an expression for $\pi_x^{(1)}$, $x \in \mathcal{T} = \{i_1, \dots, i_m\}$. Matrices A_1, S_1, and R_0 can be written as

$$A_1 = \begin{pmatrix} -s_1 & 0 \\ 0 & -s_2 \end{pmatrix},\; S_1 = \begin{pmatrix} s_1 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & s_2 \end{pmatrix},\; R_0 = \begin{pmatrix} r_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & r_2 \end{pmatrix},$$

with $s_1 = \frac{k_M^A}{V}\mathrm{D}_{\mathrm{tot}}^2$, $s_2 = \mu b \frac{k_M^A}{V}\mathrm{D}_{\mathrm{tot}}^2$, $r_1 = (k_{W0}^A + k_W^A + \frac{k_M^A}{V}(\mathrm{D}_{\mathrm{tot}} - 1))$, and $r_2 = (k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D}_{\mathrm{tot}} - 1))$. From (4.5), $\beta^{(1)} = [\pi_{i_1}^{(1)}, ..., \pi_{i_m}^{(1)}] = \alpha S_1(-T_0)^{-1}$, and so, given that the last row of T_0 is made of all zeros except for the last element, that is $(T_0)_{i_m,i_m}$ and given that $(T_0)_{i_m,i_m} = k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D}_{\mathrm{tot}} - 1)$, $\beta^{(1)} = [0, ..., 0, \pi_{i_m}^{(1)}]$, with

$$\pi_{i_m}^{(1)} = \frac{\mu b \frac{k_M^A}{V}\mathrm{D}_{\mathrm{tot}}^2}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D}_{\mathrm{tot}} - 1)}.$$

Now, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that

$$\text{(S.67)} \qquad \alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0], \qquad \alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}.$$

For an illustration, suppose $\mathrm{D}_{\mathrm{tot}} = 2$. Then $A_0 = 0$, $S_0 = 0$ and matrices $A_1 \in \mathbb{R}^{2\times 2}$ and $S_1 \in \mathbb{R}^{2\times 8}$ are given by

$$A_1 = \begin{pmatrix} -4\frac{k_M^A}{V} & 0 \\ 0 & -4\frac{k_M^A}{V}\mu b \end{pmatrix},\; S_1 = \begin{pmatrix} 4\frac{k_M^A}{V} & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & 4\frac{k_M^A}{V}\mu b \end{pmatrix}.$$

Furthermore, $R_1 = 0$ and $R_0 \in \mathbb{R}^{8\times 2}$ can be written as

$$R_0 = \begin{pmatrix} \bar{k}_W^A + \frac{k_M^A}{V} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0}^2 + \frac{k_M}{V} + \frac{\bar{k}_M}{V} \end{pmatrix}.$$

Finally, matrices T_0 and T_1 can be written as

$$T_0 = \begin{pmatrix} -\bar{k}_W^A - \frac{k_M^A}{V} - \bar{k}_W^1 & 0 & 0 & 0 & \bar{k}_W^1 & 0 & 0 & 0 \\ 0 & -\frac{k_E^A}{V}(2+\mu) & 0 & 2\frac{k_E^A}{V} & \mu\frac{k_E^A}{V} & 0 & 0 & 0 \\ 2\bar{k}_W^A & 0 & -2(\bar{k}_W^A + \bar{k}_W^1) & 0 & 0 & 2\bar{k}_W^1 & 0 & 0 \\ 0 & \bar{k}_W^A & 0 & -(\bar{k}_W^A + \bar{k}_W^1 + \frac{k_M^{'}}{V}) & 0 & 0 & 0 & \bar{k}_W^1 + \frac{k_M^{'}}{V} \\ \mu'\frac{k_E^A}{V} & k_{W0}^2 & 0 & 0 & -(\frac{k_E^A}{V}(1+\mu') + k_{W0}^2) & \frac{k_E^A}{V} & 0 & 0 \\ 0 & 0 & 0 & k_{W0}^2 & \bar{k}_W^A & -(k_{W0}^2 + \bar{k}_W^A + \bar{k}_W^1) & \bar{k}_W^1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2(k_{W0}^2 + \frac{\bar{k}_M}{V}) & 2(k_{W0}^2 + \frac{\bar{k}_M}{2V}) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2(k_{W0}^2 + \frac{k_M}{V} + \frac{\bar{k}_M}{V}) \end{pmatrix},$$

$$T_1 = \begin{pmatrix} -2\frac{k_M^A}{V} & 0 & 2\frac{k_M^A}{V} & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\frac{k_M^A}{V}(1+b\mu) & 0 & 2\frac{k_M^A}{V} & 2b\mu\frac{k_M^A}{V} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2b\mu\frac{k_M^A}{V} & 0 & 2b\mu\frac{k_M^A}{V} & 0 & 0 \\ 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & 0 & -2\frac{k_M^A}{V}(1+\beta\mu') & 2\frac{k_M^A}{V} & 0 & 0 \\ 0 & 0 & 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & -2\mu'\beta\frac{k_M^A}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4\mu'\beta\frac{k_M^A}{V} & -4\mu'\beta\frac{k_M^A}{V} & 0 \\ 0 & 0 & 0 & 2\mu'\beta\frac{k_M^A}{V} & 0 & 0 & 2\mu b\frac{k_M^A}{V} & -2(\mu'\beta\frac{k_M^A}{V} + \mu b\frac{k_M^A}{V}) \end{pmatrix},$$

∎

in which $\bar{k}_W^A = k_{W0}^A + k_W^A$, $\bar{k}_W^R = k_{W0}^R + k_W^R$. Now, by applying Theorem S.9, we first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0\ldots, 0]$ where α is the unique probability vector such that $\alpha Q_{\mathcal{A}} = 0$. In this case,

$$Q_{\mathcal{A}} = \frac{K_1 + \mu K_2}{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \tag{S.68}$$

with

$$\begin{aligned}
K_1 &= 8\bar{k}_W^1 \frac{k_M^A}{V}(\bar{k}_W^1 + \frac{k_M^{'}}{V})((\bar{k}_W^1 + k_{W0}^2)(\frac{k_E^A}{V} + k_{W0}^2) + \bar{k}_W^A k_{W0}^2), \\
K_2 &= 4\bar{k}_W^1 \frac{k_M^A}{V}\frac{k_E^A}{V}((\bar{k}_W^1 + \frac{k_M^{'}}{V})(\bar{k}_W^1 + k_{W0}^2) + \bar{k}_W^1 \bar{k}_W^A), \\
K_3 &= \frac{k_E^A}{V}(2(\bar{k}_W^1 + \bar{k}_W^A + k_{W0}^2 + \frac{k_M^{'}}{V} + \frac{k_M^A}{V}))(\bar{k}_W^1)^2 \\
&\quad + 2\bar{k}_W^1(\bar{k}_W^1(\bar{k}_W^1 + k_{W0}^2(k_{W0}^2 + \frac{k_M^{'}}{V})) + \frac{k_M^{'}}{V}(k_{W0}^2)^2) \\
&\quad + 2\frac{k_M^A}{V}k_{W0}^2(\frac{k_M^{'}}{V}(k_{W0}^2 + \bar{k}_W^A) + \bar{k}_W^1 + (2\bar{k}_W^1 + \bar{k}_W^A + k_{W0}^2)) \\
&\quad + 2\frac{k_M^A}{V}\frac{k_M^{'}}{V}k_{W0}^2(\bar{k}_W^1 + \bar{k}_W^A) + 2\bar{k}_W^1 k_{W0}^2(\frac{k_E^A}{V}\frac{k_M^{'}}{V} + \bar{k}_W^A\frac{k_M^A}{V} + 2\bar{k}_W^A\frac{k_M^{'}}{V}), \\
K_4 &= \frac{k_E^A}{V}((\bar{k}_W^1 + \frac{k_M^{'}}{V})(\bar{k}_W^A + \frac{k_M^A}{V})(\bar{k}_W^1 + k_{W0}^2) + (\bar{k}_W^1 + k_{W0}^2)\bar{k}_W^1\frac{k_M^{'}}{V}) \\
&\quad + \frac{k_E^A}{V}\bar{k}_W^1(\bar{k}_W^1(\bar{k}_W^A + k_{W0}^2 + 1) + \bar{k}_W^A(\bar{k}_W^A + \frac{k_M^A}{V})), \qquad \text{(S.69)} \\
K_5 &= 2\frac{k_E^A}{V}(\bar{k}_W^1 + k_{W0}^2 + \bar{k}_W^A)(\bar{k}_W^A + \frac{k_M^A}{V})(\bar{k}_W^1 + \frac{k_M^{'}}{V}), \\
K_6 &= \frac{k_E^A}{V}(\bar{k}_W^1 + k_{W0}^2 + \bar{k}_W^A)(\bar{k}_W^A + \frac{k_M^A}{V})(\bar{k}_W^A + \bar{k}_W^1 + \frac{k_M^{'}}{V}),
\end{aligned}$$

and then $\alpha_a = 0$ and $\alpha_r = 1$. Let us now derive an expression for $\beta^{(1)}$. Starting from the transient states $\mathcal{T} = \{i_1, \ldots, i_8\}$, we obtain that $\beta^{(1)} = [\pi^{(1)}_{i_1}, ..., \pi^{(1)}_{i_8}]$ can be determine by $\beta^{(1)} = \alpha S_1(-T_0)^{-1}$, obtaining $\beta^{(1)} = [0, ..., 0, \pi^{(1)}_{i_8}]$, with

$$\pi^{(1)}_{i_8} = \frac{4\mu b \frac{k^A_M}{V}}{k^2_{W0} + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})}.$$

Finally, $\alpha^{(1)} = [\pi^{(1)}_a, \pi^{(1)}_r]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$ and $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$. After some calculations, we obtain

$$\text{(S.70)} \qquad \pi^{(1)}_a = \frac{\mu^2\mu'^2 K_7}{K_8(K_9 + K_{10}\mu)}, \quad \pi^{(1)}_r = -\pi^{(1)}_a - \pi^{(1)}_{i_8} = -\frac{\mu^2\mu'^2 K_7 + \mu K_{11}(K_9 + K_{10}\mu)}{K_8(K_9 + K_{10}\mu)},$$

with

$$\begin{aligned}
&\bar{k}^1_W = k^1_{W0} + k^1_W, \qquad \bar{k}^A_W = k^A_{W0} + k^A_W,\\
&K_7 = 2\beta b\frac{k^A_M}{V}\bar{k}^A_W(\bar{k}^A_W + \frac{k^A_M}{V})(\bar{k}^A_W + \bar{k}^1_W + k^2_{W0}),\\
&K_8 = \bar{k}^1_W(k^2_{W0} + \frac{k_M}{V} + \frac{\bar{k}_M}{V}), \qquad K_{10} = \frac{k^A_E}{V}\left((\bar{k}^1_W + k^2_{W0})(\bar{k}^1_W + \frac{k'_M}{V}) + \bar{k}^A_W\bar{k}^1_W\right)\\
&K_9 = 2(\bar{k}^1_W + \frac{k'_M}{V})(\frac{k^A_E}{V}(\bar{k}^1_W + k^2_{W0}) + k^2_{W0}(k^2_{W0} + \bar{k}^1_W + \bar{k}^A_W)), \qquad K_{11} = 4b\bar{k}^1_W\frac{k^A_M}{V}.
\end{aligned}$$

Time to memory loss. As a reminder, we define the time to memory loss of the active state as $h_{a,r}(\varepsilon)$ and the time to memory loss of repressed state as $h_{r,a}(\varepsilon)$. Let us start by deriving the order and the leading coefficient of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$. By (4.16), we know the order of the stationary distribution at a and r are $k_a = -\min\{1-2, 0\} = 1$ and $k_r = -\min\{1-1, 0\} = 0$, respectively. This is consistent with the results in Section 5.1.1. Moreover, the leading coefficient in the stationary distribution for the fully repressed and fully active states are $\pi^{(0)}_r = 1$ and $\pi^{(1)}_a > 0$, respectively. Now, $\hat{X}_{\mathcal{A}}$ has the infinitesimal generator in the form of

$$Q_{\mathcal{A}} = \begin{pmatrix} -(Q_{\mathcal{A}})_{a,r} & (Q_{\mathcal{A}})_{a,r} \\ 0 & 0 \end{pmatrix},$$

and $\hat{X}_{\mathcal{A}}$ has a unique stationary distribution $\alpha = [0, 1]$. By Theorem 4.7,

$$D = \begin{pmatrix} \frac{1}{(Q_{\mathcal{A}})_{a,r}} & -\frac{1}{(Q_{\mathcal{A}})_{a,r}} \\ 0 & 0 \end{pmatrix},$$

$$h_{a,r}(\varepsilon) = \frac{D_{r,r} - D_{a,r}}{\pi^{(k_r)}_r}\frac{1}{\varepsilon^{k_r+1}} + O\left(\frac{1}{\varepsilon^{k_r}}\right) = \frac{D_{r,r} - D_{a,r}}{\alpha_r}\frac{1}{\varepsilon} + O(1) = \frac{1}{(Q_{\mathcal{A}})_{a,r}}\frac{1}{\varepsilon} + O(1),$$

and

$$h_{r,a}(\varepsilon) = \frac{D_{a,a} - D_{r,a}}{\pi_a^{(k_a)}} \frac{1}{\varepsilon^{k_a+1}} + O\left(\frac{1}{\varepsilon^{k_a}}\right) = \frac{D_{a,a} - D_{r,a}}{\pi_a^{(1)}} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right) = \frac{1}{(Q_{\mathcal{A}})_{a,r} \cdot \pi_a^{(1)}} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right).$$

As an example, when $\mathrm{D_{tot}} = 2$, $Q_{\mathcal{A}}$ and $\pi_a^{(1)}$ are shown in (5.3) and (5.4), and we obtain that

$$h_{a,r}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6}{K_1 + \mu K_2} \frac{1}{\varepsilon} + O(1),$$

and

$$h_{r,a}(\varepsilon) = \frac{K_3 + \mu K_4 + \mu' K_5 + \mu\mu' K_6}{K_1 + \mu K_2} \frac{K_8(K_9 + K_{10}\mu)}{\mu^2 \mu'^2 K_7} \frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right),$$

where K_i, $i = 1, ..., 11$, are non-negative functions independent of μ and μ' as defined in (S.69).

Now, let us verify that both conditions (i) and (ii) of Theorem 4.10 hold. To this end, let us first write the directions of the six possible transitions of the continuous time Markov chain $X^\varepsilon(t)$, which are $v_1 = (0,1,0)^T$, $v_2 = (0,-1,0)^T$, $v_3 = (0,0,1)^T$, $v_4 = (0,0,-1)^T$, $v_5 = (1,0,-1)^T$, $v_6 = (-1,0,1)^T$, with the associated infinitesimal transition rates that can be written as $\Upsilon_1(x) = f_A(x)$, $\Upsilon_2(x) = g_\varepsilon^A(x)$, $\Upsilon_3(x) = f_{R1}(x)$, $\Upsilon_4(x) = g_{R1}^\varepsilon(x)$, $\Upsilon_5(x) = f_{R12}(x)$, $\Upsilon_6(x) = g_{R12}^\varepsilon(x)$. Define the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and, for $x \in \mathcal{X}$, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x \preccurlyeq_A w\}$. Let us also introduce infinitesimal transition rates $\breve{\Upsilon}_i(x)$, $i = 1, 2, ..., 6$, defined as for $\Upsilon_i(x)$, $i = 1, 2, ..., 6$, with all the parameters having the same values except that μ' is replaced by $\breve{\mu}'$, with $\mu' \geq \breve{\mu}'$. Given that $Av_1 = (0,-1,0)^T$, $Av_2 = (0,1,0)^T$, $Av_3 = (0,0,1)^T$, $Av_4 = (0,0,-1)^T$, $Av_5 = (1,0,0)^T$, $Av_6 = (-1,0,0)^T$, condition (i) of Theorem 4.10 holds.

To verify condition (ii) of Theorem 4.10, consider $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_1 = w_1, x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3\} = \{w \in \mathbb{R}^3 : x_1 = w_1, x_2 \geq w_2, x_3 \leq w_3\}$. Given that $\langle A_{1\bullet}, v_5\rangle = 1$ and $\langle A_{1\bullet}, v_6\rangle = -1$, we need to verify that $\Upsilon_5(x) \leq \breve{\Upsilon}_5(y)$ and $\Upsilon_6(x) \geq \breve{\Upsilon}_6(y)$. Since $x_1 = y_1, x_2 \geq y_2, x_3 \leq y_3$, then $\Upsilon_5(x) = x_3 \left(k_{W0}^2 + \frac{k_M}{V} x_1 + \frac{\bar{k}_M}{V}(x_1 + \frac{x_3 - 1}{2})\right) \leq y_3 \left(k_{W0}^2 + \frac{k_M}{V} x_1 + \frac{\bar{k}_M}{V}(y_1 + \frac{y_3 - 1}{2})\right) = \breve{\Upsilon}_5(y)$ and $\Upsilon_6(x) = x_1 \mu \left(\varepsilon \frac{k_M^A}{V} \mathrm{D_{tot}} b + x_2 \frac{k_E^A}{V}\right)$ $\geq y_1 \mu \left(\varepsilon \frac{k_M^A}{V} \mathrm{D_{tot}} b + y_2 \frac{k_E^A}{V}\right) = \breve{\Upsilon}_6(y)$. Let us now consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : x_1 \leq w_1, x_2 = w_2, x_1 + x_3 \leq w_1 + w_3\}$. Given that $\langle A_{2\bullet}, v_1\rangle = -1$ and $\langle A_{2\bullet}, v_2\rangle = 1$, we need to verify that $\Upsilon_1(x) \geq \breve{\Upsilon}_1(y)$ and $\Upsilon_2(x) \leq \breve{\Upsilon}_2(y)$. Since $x_1 \leq y_1, x_2 = y_2, x_1 + x_3 \leq y_1 + y_3$, then $\Upsilon_1(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} x_2\right) \geq (\mathrm{D_{tot}} - (y_1 + y_2 + y_3)) \left(k_{W0}^A + k_W^A + \frac{k_M^A}{V} y_2\right) = \breve{\Upsilon}_1(y)$ and $\Upsilon_2(x) = x_2 \left(\varepsilon \frac{k_M^A}{V} \mathrm{D_{tot}} + \frac{k_E^A}{V}(x_3 + 2x_1)\right) \leq y_2 \left(\varepsilon \frac{k_M^A}{V} \mathrm{D_{tot}} + \frac{k_E^A}{V}(y_3 + 2y_1)\right) = \breve{\Upsilon}_2(y)$. Finally, consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in$

$\mathcal{X} \mid x_1 \le w_1, x_2 \ge w_2, x_1 + x_3 = w_1 + w_3\} = \{w \in \mathcal{X} \mid x_1 \le w_1, x_2 \ge w_2, x_3 \ge w_3\}$. Given that $\langle A_{3\bullet}, v_3 \rangle = 1$ and $\langle A_{3\bullet}, v_4 \rangle = -1$, we need to check that $\Upsilon_3(x) \le \breve{\Upsilon}_3(y)$ and $\Upsilon_4(x) \ge \breve{\Upsilon}_4(y)$. Since $x_1 \le y_1, x_2 \ge y_2, x_3 \ge y_3$, then $\Upsilon_3(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3)) \left(k^1_{W0} + k^1_W + \frac{k'_M}{V} x_1 \right) \le (\mathrm{D_{tot}} - (y_1 + y_2 + y_3)) \left(k^1_{W0} + k^1_W + \frac{k'_M}{V} y_1 \right) = \breve{\Upsilon}_3(y)$ and $\Upsilon_4(x) = x_3 \mu' \left(\varepsilon \frac{k^A_M}{V} \mathrm{D_{tot}} \beta + x_2 \frac{k^A_E}{V} \right) \ge y_3 \breve{\mu}' \left(\varepsilon \frac{k^A_M}{V} \mathrm{D_{tot}} \beta + y_2 \frac{k^A_E}{V} \right) = \breve{\Upsilon}_4(y)$.

We can then conclude that all of the conditions of Theorem 4.10 hold.

S.11 4D Model: additional mathematical details

Verification of Assumption 4.1. In order to show that Assumption 4.1 holds, consider the states $a = (0, \mathrm{D_{tot}}, 0, 0)^T$ and $r = (\mathrm{D_{tot}}, 0, 0, 0)^T$ and the set $\mathcal{T} = \{i_1, \ldots, i_m\}$ defined in Section 5.2.1. From (5.9), we can see that $Q_{a,a+v_j}(0) = Q_{r,r+v_j}(0) = 0$ for every $1 \le j \le 10$. As a consequence, both a and r are absorbing states under $Q(0)$. To see that the states in $\mathcal{T}$ are transient under $Q(0)$, consider a state $x = (x_1, x_2, x_3, x_4)^T \in \mathcal{T}$. First, suppose $x_1 + x_3 + x_4 \ne 0$. By having the one-step transitions along $v_2 = (0, -1, 0, 0)^T$ occurring x_2 times where $Q_{z,z+v_2}(0) = \frac{k^A_E}{V}(x_3 + x_4 + 2x_1) z_2 > 0$ for all $z = (x_1, z_2, x_3, x_4)^T$ and $1 \le z_2 \le x_2$, then having one-step transitions along $v_3 = (0, 0, 1, 0)^T$ occurring $\mathrm{D_{tot}} - x_1 - x_3 - x_4$ times where $Q_{z,z+v_3}(0) = (\mathrm{D_{tot}} - (x_1 + z_3 + x_4)) \left(k^1_{W0} + k^1_W + \frac{k'_M}{V}(x_1 + x_4) \right) > 0$ for all $z = (x_1, 0, z_3, x_4)^T$ and $x_3 \le z_3 \le \mathrm{D_{tot}} - x_1 - x_4 - 1$, then having one-step transitions along $v_9 = (1, 0, 0, -1)^T$ occurring x_4 times where $Q_{z,z+v_9}(0) = (x_1 + x_4 - z_1) \left(k^1_{W0} + \frac{k'_M}{V} \frac{x_1 + x_4 + z_1 - 1}{2} \right) > 0$ for all $z = (z_1, 0, \mathrm{D_{tot}} - x_1 - x_4, x_1 + x_4 - z_1)^T$ and $x_1 \le z_1 \le x_1 + x_4 - 1$, and finally having one-step transitions along $v_7 = (1, 0, -1, 0)^T$ occurring $\mathrm{D_{tot}} - x_1 - x_4$ times where $Q_{z,z+v_7}(0) = (\mathrm{D_{tot}} - z_1) \left(k^2_{W0} + \frac{k_M}{V} z_1 + \frac{\bar{k}_M}{V} \frac{\mathrm{D_{tot}} + z_1 - 1}{2} \right) > 0$ for all $z = (z_1, 0, \mathrm{D_{tot}} - z_1, 0)^T$ and $x_1 + x_4 \le z_1 \le \mathrm{D_{tot}} - 1$, we have a positive probability of transition from x to r under $Q(0)$. By Lemma S.8 and the fact that r is an absorbing state, we have that x is a transient state for X^0. On the other hand, suppose $x_1 + x_3 + x_4 = 0$. Since $x = (0, x_2, 0, 0)^T \in \mathcal{T}$, we have $0 \le x_2 \le \mathrm{D_{tot}} - 1$. We can first have a one-step transition along $v_3 = (0, 0, 1, 0)^T$, where $Q_{x,x+v_3}(0) = (\mathrm{D_{tot}} - x_2)(k^1_{W0} + k^1_W) > 0$, to reach the state $(0, x_2, 1, 0)^T$ and then take the steps in the $x_1 + x_3 + x_4 \ne 0$ case. In this way, there is a positive probability of transition from x to the absorbing state r, and thus x is transient by Lemma S.8.

Verification of Assumption 4.2. To show that Assumption 4.2 holds, consider the continuous time Markov chain $\tilde{X}$ with infinitesimal generator $\tilde{Q}$ as described in (4.7) and shown in Fig. 7(d). We will first see that $\{i_{m-1}, i_m, r\}$, with $i_{m-1} = r + v_{10}$ and $i_m = r + v_8$, forms a closed class under $\tilde{Q}$. For this, we see that $Q_{r,r+v_j}(\varepsilon)$ vanishes for every $j = \{1, 2, 3, 4, 5, 6, 7, 9\}$ and $\varepsilon \ge 0$, while $Q_{r,r+v_8}(\varepsilon) = \varepsilon \mu b \frac{k^A_M}{V} \mathrm{D^2_{tot}}$ and $Q_{r,r+v_{10}}(\varepsilon) = \varepsilon \mu' \beta \frac{k^A_M}{V} \mathrm{D^2_{tot}}$. Therefore, the only transitions from r under $\tilde{Q}$ are to i_{m-1} with rate $\tilde{Q}_{r,i_{m-1}} = \mu' \beta \frac{k^A_M}{V} \mathrm{D^2_{tot}} > 0$ and to i_m with rate $\tilde{Q}_{r,i_m} = \mu b \frac{k^A_M}{V} \mathrm{D^2_{tot}} > 0$. From (5.9), we can see that $Q_{i_{m-1},i_{m-1}+v_j}(0) = 0$ for $j \in \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$, $Q_{i_m,i_m+v_j}(0) = 0$ for $j \in \{1, 2, 3, 4, 5, 6, 8, 9, 10\}$, $Q_{i_{m-1},r}(0) = Q_{i_{m-1},i_{m-1}+v_9}(0) = k^1_{W0} + \frac{k'_M}{V}(\mathrm{D_{tot}} - 1) > 0$ and $Q_{i_m,r}(0) = Q_{i_m,i_m+v_7}(0) = k^2_{W0} + \frac{k_M}{V}(\mathrm{D_{tot}} -$

$1) + \frac{\bar{k}_M}{V}(\mathrm{D_{tot}} - 1) > 0$. Therefore, $\{i_{m-1}, i_m, r\}$ forms a closed class under $\tilde{Q}$. The fact that $\hat{X}_{\mathcal{A}}$ consists of erasing the times from $\tilde{X}$ in which the process is in $\mathcal{T}$, together with Lemma 4.3, yields that r is an absorbing state under $Q_{\mathcal{A}}$. From (5.9), we can see that $\tilde{Q}_{a,i_1} > 0$ where $i_1 = (0, \mathrm{D_{tot}} - 1, 0, 0)^T$. From the analysis made to prove Assumption 4.1, we obtain that i_1 leads to r under $\tilde{Q}$, which is part of the closed class $\{i_{m-1}, i_m, r\}$. By interpreting $\hat{X}_{\mathcal{A}}$ again as a time-change of $\tilde{X}$, by Lemma 4.3 we obtain that a is transient under $Q_{\mathcal{A}}$. As a consequence, $Q_{\mathcal{A}}$ has a single recurrent class consisting of the state r. Thus, Assumption 4.2 holds, and furthermore, $\alpha = [\alpha_a, \alpha_r]$ with $\alpha_a = 0$ and $\alpha_r = 1$. In addition, the previous arguments show that neither Assumption 4.3 nor 4.4 holds for this model.

Stationary distribution. Here, we derive an expression for $\pi_x^{(1)}$, $x \in \mathcal{T} = \{i_1, \ldots, i_m\}$. Matrices A_1, S_1, and R_0 can be written as

$$A_1 = \begin{pmatrix} -s_1 & 0 \\ 0 & -(s_2 + s_3) \end{pmatrix}, \; S_1 = \begin{pmatrix} s_1 & 0 & \ldots & \ldots & 0 & 0 \\ 0 & \ldots & \ldots & 0 & s_2 & s_3 \end{pmatrix}, \; R_0 = \begin{pmatrix} r_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & r_2 \\ 0 & r_3 \end{pmatrix},$$

with $s_1 = \frac{k_M^A}{V}\mathrm{D_{tot}^2}$, $s_2 = \mu'\beta\varepsilon\frac{k_M^A}{V}\mathrm{D_{tot}^2}$, $s_3 = \mu b\varepsilon\frac{k_M^A}{V}\mathrm{D_{tot}^2}$, $r_1 = (k_{W0}^A + k_W^A + \frac{k_M^A}{V}(\mathrm{D_{tot}} - 1))$, $r_2 = (k_{W0}^1 + \frac{k_M'}{V}(\mathrm{D_{tot}} - 1))$ and $r_3 = (k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D_{tot}} - 1))$. Now, we determine $\beta^{(1)} = [\pi_{i_1}^{(1)}, ..., \pi_{i_m}^{(1)}] = \alpha S_1(-T_0)^{-1}$. Given that the only two elements different from zero in the last two rows of T_0 are $(T_0)_{i_{m-1}, i_{m-1}} = (k_{W0}^1 + \frac{k_M'}{V}(\mathrm{D_{tot}} - 1)$ and $(T_0)_{i_m, i_m} = (k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V}(\mathrm{D_{tot}} - 1))$, we obtain $\beta^{(1)} = [0, ..., 0, \pi_{i_{m-1}}^{(1)}, \pi_{i_m}^{(1)}]$, with

$$\pi_{i_{m-1}}^{(1)} = \frac{\mu'\beta\frac{k_M^A}{V}\mathrm{D_{tot}^2}}{k_{W0}^1 + \frac{k_M'}{V}(\mathrm{D_{tot}} - 1)}, \qquad \pi_{i_m}^{(1)} = \frac{\mu b\frac{k_M^A}{V}\mathrm{D_{tot}^2}}{k_{W0}^2 + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})(\mathrm{D_{tot}} - 1)}.$$

Now, $\alpha^{(1)} = [\pi_a^{(1)}, \pi_r^{(1)}]$ is the unique vector such that $\alpha^{(1)} Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$, $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$.

As an example, suppose $\mathrm{D_{tot}} = 2$, $\beta = b$, $k_W^1 = k_W^2 = k_W^A = 0$, $k_{W0}^1 = k_{W0}^2 = k_{W0}^A = k_{W0}$ and $k_M' = \bar{k}_M = k_M^A = k_M$. Then, we have that $A_0 = 0$, $S_0 = 0$ and matrices $A_1 \in \mathbb{R}^{2\times 2}$ and $S_1 \in \mathbb{R}^{2\times 13}$ are equal to

$$A_1 = \begin{pmatrix} -4\frac{k_M}{V} & 0 \\ 0 & -4\frac{k_M}{V}(\mu b + \mu'\beta) \end{pmatrix}, \; S_1 = \begin{pmatrix} 4\frac{k_M}{V} & 0 & \ldots & \ldots & 0 & 0 \\ 0 & \ldots & \ldots & 0 & 4\frac{k_M}{V}\mu' b & 4\frac{k_M}{V}\mu b \end{pmatrix}.$$

Furthermore, $R_1 = 0$ and $R_0 \in \mathbb{R}^{13\times 2}$ can be written as

$$R_0 = \begin{pmatrix} k_{W0} + \frac{k_M}{V} & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & k_{W0} + \frac{k_M}{V} \\ 0 & k_{W0} + 2\frac{k_M}{V} \end{pmatrix}.$$

Finally, matrices T_0 and T_1 can be written as

$$T_0 = \begin{pmatrix} T_0^1 & T_0^2 \\ T_0^3 & T_0^4 \end{pmatrix}, T_1 = \begin{pmatrix} T_1^1 & T_1^2 \\ T_1^3 & T_1^4 \end{pmatrix},$$

with $T_1^2 = 0^{7\times 6}$ and

$$T_0^1 = \begin{pmatrix} -3k_{W0} - \frac{k_M}{V} & k_{W0} & 0 & 0 & 0 & k_{W0} & 0 \\ \mu\frac{k_E^A}{V} & -(1+\mu)\frac{k_E^A}{V} - k_{W0} & 0 & \frac{k_E^A}{V} & 0 & 0 & 0 \\ 2k_{W0} & 0 & -6k_{W0} & 2k_{W0} & 0 & 0 & 2k_{W0} \\ 0 & k_{W0} & 0 & -(4k_{W0} + 2\frac{k_M}{V}) & k_{W0} + \frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & -(k_{W0} + \frac{k_M}{V}) & 0 & 0 \\ \mu'\frac{k_E^A}{V} & 0 & 0 & 0 & 0 & -(1+\mu')\frac{k_E^A}{V} - k_{W0} & \frac{k_E^A}{V} \\ 0 & 0 & 0 & 0 & 0 & k_{W0} & -(4k_{W0} + \frac{k_M}{V}) \end{pmatrix},$$

$$T_0^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{W0} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ k_{W0} + \frac{k_M}{V} & 0 & 0 & k_{W0} & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} & 0 \\ 0 & 0 & k_{W0} & 0 & 0 & 0 \\ k_{W0} + \frac{k_M}{V} & k_{W0} & 0 & k_{W0} & 0 & 0 \end{pmatrix}, \quad T_1^3 = \begin{pmatrix} 0 & 0 & 0 & 2\mu' b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\mu' b\frac{k_M}{V} \\ 0 & 2\mu' b\frac{k_M}{V} & 0 & 0 & 0 & 2\mu b\frac{k_M}{V} & 0 \\ 0 & 0 & 0 & 2\mu' b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} \\ 0 & 0 & 0 & 0 & 2\mu' b\frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T_1^1 = \begin{pmatrix} -2\frac{k_M}{V} & 0 & 2\frac{k_M}{V} & 0 & 0 & 0 & 0 \\ 2\mu b\frac{k_M}{V} & -2(\mu b + 1)\frac{k_M}{V} & 0 & 2\frac{k_M}{V} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\mu b\frac{k_M}{V} & -2\mu b\frac{k_M}{V} & 0 & 0 & 0 \\ 2\mu' b\frac{k_M}{V} & 0 & 0 & 0 & 0 & -2(\mu' b + 1)\frac{k_M}{V} & 2\frac{k_M}{V} \\ 0 & 0 & 2\mu' b\frac{k_M}{V} & 0 & 0 & 0 & -2\mu' b\frac{k_M}{V} \end{pmatrix}, \quad T_0^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu'\frac{k_E^A}{V} & 0 & 0 & 0 & \mu\frac{k_E^A}{V} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T_0^4 = \begin{pmatrix} -(2k_{W0} + \frac{k_M}{V}) & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} & k_{W0} \\ 0 & -(k_{W0} + \frac{k_M}{V}) & 0 & 0 & 0 & k_{W0} + \frac{k_M}{V} \\ 0 & 0 & -(\mu + \mu' + 2)\frac{k_E^A}{V} & 2\frac{k_E^A}{V} & 0 & 0 \\ 0 & 0 & k_{W0} & -3(k_{W0} + \frac{k_M}{V}) & k_{W0} + 2\frac{k_M}{V} & k_{W0} + \frac{k_M}{V} \\ 0 & 0 & 0 & 0 & -(k_{W0} + \frac{k_M}{V}) & 0 \\ 0 & 0 & 0 & 0 & 0 & -(k_{W0} + 2\frac{k_M}{V}) \end{pmatrix},$$

$$T_1^4 = \begin{pmatrix} -2(\mu + \mu')b\frac{k_M}{V} & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\mu' b\frac{k_M}{V} & 0 & 0 & 0 & 0 \\ 0 & 0 & -2(\mu b + \mu' b + 1)\frac{k_M}{V} & 2\frac{k_M}{V} & 0 & 0 \\ 0 & 0 & 0 & -2(\mu + \mu')b\frac{k_M}{V} & 0 & 0 \\ 2\mu b\frac{k_M}{V} & 0 & 0 & 2\mu b\frac{k_M}{V} & -2(2\mu + \mu')b\frac{k_M}{V} & 0 \\ 2\mu' b\frac{k_M}{V} & 2\mu b\frac{k_M}{V} & 0 & 2\mu' b\frac{k_M}{V} & 0 & -2(\mu + 2\mu')b\frac{k_M}{V} \end{pmatrix},$$ ∎

Now, by applying Theorem S.9, we first obtain that $\pi(0) = \pi^{(0)} = [\alpha, 0] = [\alpha_a, \alpha_r, 0\ldots, 0]$

where α can be obtained by solving $\alpha Q_{\mathcal{A}} = 0$. In this case, we obtain that

$$Q_{\mathcal{A}} = \frac{K_1((\mu')^2K_2 + (\mu)^2K_3 + \mu\mu'K_4 + \mu'K_5 + \mu K_6 + K_7)}{\mu'\mu(\mu'+\mu)K_8 + (\mu')^2K_9 + (\mu)^2K_{10} + \mu\mu'K_{11} + \mu'K_{12} + \mu K_{13} + K_{14}} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix},$$

with

$$\begin{aligned}
K_1 &= 4k_{W0}\frac{k_M}{V},\\
K_2 &= \left(\frac{k_E^A}{V}\right)^2\left(6\left(\frac{k_M}{V}\right)^3 + 39k_{W0}\left(\frac{k_M}{V}\right)^2 + 68\frac{k_M}{V}(k_{W0})^2 + 32(k_{W0})^3\right),\\
K_3 &= \left(\frac{k_E^A}{V}\right)^2\left(6\left(\frac{k_M}{V}\right)^3 + 36k_{W0}\left(\frac{k_M}{V}\right)^2 + 64\frac{k_M}{V}(k_{W0})^2 + 32(k_{W0})^3\right),\\
K_4 &= \left(\frac{k_E^A}{V}\right)^2\left(12\left(\frac{k_M}{V}\right)^3 + 75k_{W0}\left(\frac{k_M}{V}\right)^2 + 132\frac{k_M}{V}(k_{W0})^2 + 64(k_{W0})^3\right),\\
K_5 &= \left(\frac{k_E^A}{V}\right)^2\left(24\left(\frac{k_M}{V}\right)^3 + 140k_{W0}\left(\frac{k_M}{V}\right)^2 + 222\frac{k_M}{V}(k_{W0})^2 + 96(k_{W0})^3\right)\\
&\quad + \frac{k_E^A}{V}k_{W0}\left(24\left(\frac{k_M}{V}\right)^3 + 158k_{W0}\left(\frac{k_M}{V}\right)^2 + 280\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right),\\
K_6 &= \left(\frac{k_E^A}{V}\right)^2\left(24\left(\frac{k_M}{V}\right)^3 + 134k_{W0}\left(\frac{k_M}{V}\right)^2 + 216\frac{k_M}{V}(k_{W0})^2 + 96(k_{W0})^3\right)\\
&\quad + \frac{k_E^A}{V}k_{W0}\left(24\left(\frac{k_M}{V}\right)^3 + 152k_{W0}\left(\frac{k_M}{V}\right)^2 + 272\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right),\\
K_7 &= \left(\frac{k_E^A}{V}\right)^2\left(24\left(\frac{k_M}{V}\right)^3 + 124k_{W0}\left(\frac{k_M}{V}\right)^2 + 180\frac{k_M}{V}(k_{W0})^2 + 72(k_{W0})^3\right)\\
&\quad + \frac{k_E^A}{V}k_{W0}\left(48\left(\frac{k_M}{V}\right)^3 + 284k_{W0}\left(\frac{k_M}{V}\right)^2 + 456\frac{k_M}{V}(k_{W0})^2 + 192(k_{W0})^3\right)\\
&\quad + (k_{W0})^2\left(24\left(\frac{k_M}{V}\right)^3 + 160k_{W0}\left(\frac{k_M}{V}\right)^2 + 288\frac{k_M}{V}(k_{W0})^2 + 128(k_{W0})^3\right),\\
K_8 &= \left(\frac{k_E^A}{V}\right)^2\left(6\left(\frac{k_M}{V}\right)^4 + 48k_{W0}\left(\frac{k_M}{V}\right)^3 + 126(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 132\frac{k_M}{V}(k_{W0})^3 + 72(k_{W0})^4\right),\\
K_9 &= \left(\frac{k_E^A}{V}\right)^2\left(6\left(\frac{k_M}{V}\right)^4 + 51k_{W0}\left(\frac{k_M}{V}\right)^3 + 146(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 168\frac{k_M}{V}(k_{W0})^3 + 64(k_{W0})^4\right),\\
K_{10} &= \left(\frac{k_E^A}{V}\right)^2\left(6\left(\frac{k_M}{V}\right)^4 + 48k_{W0}\left(\frac{k_M}{V}\right)^3 + 136(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 160\frac{k_M}{V}(k_{W0})^3 + 64(k_{W0})^4\right),\\
K_{11} &= \left(\frac{k_E^A}{V}\right)^2\left(24\left(\frac{k_M}{V}\right)^4 + 191k_{W0}\left(\frac{k_M}{V}\right)^3 + 509(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 547\frac{k_M}{V}(k_{W0})^3 + 200(k_{W0})^4\right),\\
&\quad + \frac{k_E^A}{V}k_{W0}\left(12\left(\frac{k_M}{V}\right)^4 + 96k_{W0}\left(\frac{k_M}{V}\right)^3 + 252(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 264\frac{k_M}{V}(k_{W0})^3 + 96(k_{W0})^4\right),\\
K_{12} &= \left(\frac{k_E^A}{V}\right)^2\left(19\left(\frac{k_M}{V}\right)^4 + 149k_{W0}\left(\frac{k_M}{V}\right)^3 + 416(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 463\frac{k_M}{V}(k_{W0})^3 + 168(k_{W0})^4\right),\\
&\quad + \frac{k_E^A}{V}k_{W0}\left(18\left(\frac{k_M}{V}\right)^4 + 161k_{W0}\left(\frac{k_M}{V}\right)^3 + 489(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 588\frac{k_M}{V}(k_{W0})^3 + 224(k_{W0})^4\right),
\end{aligned}$$

$$K_{13} = \left(\frac{k_E^A}{V}\right)^2 \left(18\left(\frac{k_M}{V}\right)^4 + 143k_{W0}\left(\frac{k_M}{V}\right)^3 + 399(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 452\frac{k_M}{V}(k_{W0})^3 + 168(k_{W0})^4\right),$$
$$+ \frac{k_E^A}{V}k_{W0}\left(18\left(\frac{k_M}{V}\right)^4 + 158k_{W0}\left(\frac{k_M}{V}\right)^3 + 476(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 576\frac{k_M}{V}(k_{W0})^3 + 224(k_{W0})^4\right),$$
$$K_{14} = \left(\frac{k_E^A}{V}\right)^2 \left(12\left(\frac{k_M}{V}\right)^4 + 98k_{W0}\left(\frac{k_M}{V}\right)^3 + 276(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 306\frac{k_M}{V}(k_{W0})^3 + 108(k_{W0})^4\right),$$
$$+ \frac{k_E^A}{V}k_{W0}\left(24\left(\frac{k_M}{V}\right)^4 + 214k_{W0}\left(\frac{k_M}{V}\right)^3 + 654(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 780\frac{k_M}{V}(k_{W0})^3 + 288(k_{W0})^4\right),$$
$$+ (k_{W0})^2\left(12\left(\frac{k_M}{V}\right)^4 + 116k_{W0}\left(\frac{k_M}{V}\right)^3 + 384(k_{W0})^2\left(\frac{k_M}{V}\right)^2 + 496\frac{k_M}{V}(k_{W0})^3 + 192(k_{W0})^4\right).$$

Let us now derive an expression for $\pi^{(1)}$. Starting with the transient states $\mathcal{T} = \{i_1, \ldots, i_{15}\}$, we obtain that $\beta^{(1)} = [\pi^{(1)}_{i_1}, ..., \pi^{(1)}_{i_{15}}] = \alpha S_1(-T_0)^{-1}$, and so $\beta^{(1)} = [0, ..., 0, \pi^{(1)}_{i_{14}}, \pi^{(1)}_{i_{15}}]$, with

$$\pi^{(1)}_{i_{12}} = \frac{4\mu'\beta\frac{k_M^A}{V}}{k^1_{W0} + \frac{k'_M}{V}}, \qquad \pi^{(1)}_{i_{13}} = \frac{4\mu b\frac{k_M^A}{V}}{k^2_{W0} + (\frac{k_M}{V} + \frac{\bar{k}_M}{V})}.$$

Finally, $\alpha^{(1)} = [\pi^{(1)}_a, \pi^{(1)}_r]$ is the unique vector such that $\alpha^{(1)}Q_{\mathcal{A}} = -\beta^{(1)}[R_1 + T_1(-T_0)^{-1}R_0]$ and $\alpha^{(1)}\mathbb{1} = -\beta^{(1)}\mathbb{1}$. After some calculations, we obtain

$$\begin{aligned}
\pi^{(1)}_a &= \frac{(\mu\mu')^2K_{15}((\mu+\mu')K_{16}+K_{17})}{K_{20}((\mu')^2K_2 + (\mu)^2K_3 + (\mu'+\mu)K_4 + \mu'K_5 + \mu K_6 + K_7)},\\
\pi^{(1)}_r &= -\pi^{(1)}_a - \pi^{(1)}_{i_{14}} - \pi^{(1)}_{i_{15}}\\
&= -\frac{(\mu\mu')^2K_{15}((\mu+\mu')K_{16}+K_{17})}{K_{20}((\mu')^2K_2 + (\mu)^2K_3 + (\mu'+\mu)K_4 + \mu'K_5 + \mu K_6 + K_7)} - \mu'K_{18} - \mu K_{19},
\end{aligned}$$

with

$$K_{15} = 2\frac{k_E^A}{V}\frac{k_M}{V}\left(3\frac{k_M}{V} + 2k_{W0}\right)b^2, \quad K_{16} = \frac{k_E^A}{V}\left(2\left(\frac{k_M}{V}\right)^2 + 16(k_{W0})^2 + 12\frac{k_M}{V}k_{W0}\right),$$
$$K_{17} = \frac{k_E^A}{V}\left(21\frac{k_M}{V}k_{W0} + 4\left(\frac{k_M}{V}\right)^2 + 24(k_{W0})^2\right) + k_{W0}\left(24\frac{k_M}{V}k_{W0} + 4\left(\frac{k_M}{V}\right)^2 + 32(k_{W0})^2\right),$$
$$K_{18} = \frac{4b\frac{k_M}{V}}{k_W + \frac{k_M}{V}}, \quad K_{19} = \frac{4b\frac{k_M}{V}}{k_W + 2\frac{k_M}{V}}, K_{20} = \left(2\frac{k_M}{V} + k_{W0}\right).$$

Time to memory loss. As a reminder, we define the time to memory loss of the active state as $h_{a,r}(\varepsilon)$ and the time to memory loss of the repressed state as $h_{r,a}(\varepsilon)$. Let us start by deriving the order and the leading coefficients of $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$. By (4.16), the order of the stationary distribution at a and r are $k_a = -\min\{1-2, 0\} = 1$ and $k_r = -\min\{1-1, 0\} = 0$, respectively. This is consistent with the results obtained in Section 5.2.1. As obtained for the 3D model, here we obtain $\pi^{(0)}_r = 1$ and $\pi^{(1)}_a > 0$, and thus

$$h_{a,r}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r}}\frac{1}{\varepsilon} + O(1), \text{ and } h_{r,a}(\varepsilon) = \frac{1}{(Q_{\mathcal{A}})_{a,r}\cdot\pi^{(1)}_a}\frac{1}{\varepsilon^2} + O\left(\frac{1}{\varepsilon}\right).$$

Now, in order to exploit Theorems S.2 and 3.4 from [13] and determine how μ' affects $h_{a,r}(\varepsilon)$ and $h_{r,a}(\varepsilon)$, we introduce a small approximation in the transition rates of X^ε, namely, $\frac{x_3-1}{2} \approx x_3$ and $\frac{x_4-1}{2} \approx x_4$ in $f_{R121}(x)$ and $f_{R122}(x)$, respectively. This approximation can be justified by introducing the reasonable assumption that each nucleosome characterized by a repressive modification (D_1^R and D_2^R) has the ability to catalyze the establishment of the opposite repressive mark on itself. Now, let us verify that both conditions (i) and (ii) of Theorem S.2 in [5] hold. These conditions can be written as follows:

(i) For each $1 \le j \le n$, the vector Av_j has entries in $\{-1, 0, 1\}$ only.

(ii) For each $x \in \mathcal{X}$, $1 \le i \le m$ and $y \in \partial_i(K_A + x) \cap \mathcal{X}$ we have that for each $1 \le k \le s$,

$$\sum_{j \in G_i^{k,-}} \breve{\Upsilon}_j(y) \le \sum_{j \in G_i^{k,-}} \Upsilon_j(x), \qquad \text{where } G_i^{k,-} = \{j \in G^k \mid \langle A_{i\bullet}, v_j\rangle = -1\},$$

and

$$\sum_{j \in G_i^{k,+}} \breve{\Upsilon}_j(y) \ge \sum_{j \in G_i^{k,+}} \Upsilon_j(x), \qquad \text{where } G_i^{k,+} = \{j \in G^k \mid \langle A_{i\bullet}, v_j\rangle = 1\}.$$

To verify that these conditions hold, let us first note the ten possible transitions vectors for the continuous time Markov chain $X^\varepsilon(t)$: $v_1 = -v_2 = (1, 0, -1, 0)^T$, $v_3 = -v_4 = (1, 0, 0, -1)^T$, $v_5 = -v_6 = (0, 1, 0, 0)^T$, $v_7 = -v_8 = (0, 0, 1, 0)^T$, $v_9 = -v_{10} = (0, 0, 01)^T$, with the associated infinitesimal transition rates $\Upsilon_1(x) = f_{R121}(x)$, $\Upsilon_2(x) = g^\varepsilon_{R121}(x)$, $\Upsilon_3(x) = f_{R122}(x)$, $\Upsilon_4(x) = g^\varepsilon_{R122}(x)$, $\Upsilon_5(x) = f_A(x)$, $\Upsilon_6(x) = g^\varepsilon_A(x)$, $\Upsilon_7(x) = f_{R1}(x)$, $\Upsilon_8(x) = g^\varepsilon_{R1}(x)$, $\Upsilon_9(x) = f_{R2}(x)$, $\Upsilon_{10}(x) = g^\varepsilon_{R2}(x)$. Let

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Then, $(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \ x \preccurlyeq_A w\}$. Consider infinitesimal transition rates $\breve{\Upsilon}_i(x)$, $i = 1, 2, ..., 10$, defined as for $\Upsilon_i(x)$, $i = 1, 2, ..., 10$, with all the parameters having the same values except that μ' is replaced by $\breve{\mu}'$, with $\mu' \ge \breve{\mu}'$. Now, condition (i) of Theorem S.2 in [5] holds since $Av_1 = -Av_2 = (0, 0, 1, 0)^T$, $Av_3 = -Av_4 = (0, 1, 0, 0)^T$, $Av_5 = -Av_6 = (-1, 0, 0, 0)^T$, $Av_7 = -Av_8 = (0, 1, 0, 1)^T$ and $Av_9 = -Av_{10} = (0, 0, 1, 1)^T$. Assumption S.1 in [5] holds with $G^1 = \{9, 1\}$, $G^2 = \{10, 2\}$, $G^3 = \{7, 3\}$, $G^4 = \{8, 4\}$, $G^5 = \{5\}$, $G^6 = \{6\}$ and $\sigma(1) = 9$, $\sigma(2) = 1$, $\sigma(3) = 10$, $\sigma(4) = 2$, $\sigma(5) = 7$, $\sigma(6) = 3$, $\sigma(7) = 8$, $\sigma(8) = 4$, $\sigma(9) = 5$, $\sigma(10) = 6$. To verify that also condition (ii) of Theorem S.2 in [5] holds, let us start with considering $x \in \mathcal{X}$ and $y \in \partial_1(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \ x_2 = w_2, x_1 + x_3 \le w_1 + w_3, x_1 + x_4 \le w_1 + w_4, x_1 + x_3 + x_4 \le w_1 + w_3 + w_4\}$. Given that $\langle A_{1\bullet}, v_5\rangle = -1$ and $\langle A_{1\bullet}, v_6\rangle = 1$, we must verify that $\Upsilon_5(x) \ge \breve{\Upsilon}_5(y)$ and $\Upsilon_6(x) \le \breve{\Upsilon}_6(y)$. Since $x_2 = y_2, x_1 + x_3 \le y_1 + y_3, x_1 + x_4 \le y_1 + y_4, x_1 + x_3 + x_4 \le y_1 + y_3 + y_4$, then $\Upsilon_5(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3 + x_4))\left(k^A_{W0} + k^A_W + \frac{k^A_M}{V} x_2\right) \ge (\mathrm{D_{tot}} - (y_1 + y_2 + y_3 + y_4))\left(k^A_{W0} + k^A_W + \frac{k^A_M}{V} y_2\right) = \breve{\Upsilon}_5(y)$ and $\Upsilon_6(x) = x_2\left(\varepsilon \frac{k^A_M}{V} \mathrm{D_{tot}} + \frac{k^A_E}{V}(x_3 + x_4 + 2x_1)\right) \le y_2\left(\varepsilon \frac{k^A_M}{V} \mathrm{D_{tot}} + \frac{k^A_E}{V}(y_3 + y_4 + 2y_1)\right) = \breve{\Upsilon}_6(y)$.

Let us now consider $x \in \mathcal{X}$ and $y \in \partial_2(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \; x_2 \geq w_2, x_1 + x_3 = w_1 + w_3, x_1 + x_4 \leq w_1 + w_4, x_1 + x_3 + x_4 \leq w_1 + w_3 + w_4\}$. Given that $\langle A_{2\bullet}, v_3\rangle = \langle A_{2\bullet}, v_7\rangle = 1$ and $\langle A_{2\bullet}, v_4\rangle = \langle A_{2\bullet}, v_8\rangle = -1$, we need to verify that $\Upsilon_3(x) + \Upsilon_7(x) \leq \breve{\Upsilon}_3(y) + \breve{\Upsilon}_7(y)$ and $\Upsilon_4(x) + \Upsilon_8(x) \geq \breve{\Upsilon}_4(y) + \breve{\Upsilon}_8(y)$ hold. Since $x_2 \geq y_2, x_1 + x_3 = y_1 + y_3, x_1 + x_4 \leq y_1 + y_4, x_1 + x_3 + x_4 \leq y_1 + y_3 + y_4$, then $\Upsilon_3(x) + \Upsilon_7(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3))\left(k^1_{W0} + \frac{k'_M}{V}(x_1 + x_4)\right) \leq (\mathrm{D_{tot}} - (y_1 + y_2 + y_3))\left(k^1_{W0} + \frac{k'_M}{V}(y_1 + y_4)\right) = \breve{\Upsilon}_3(y) + \breve{\Upsilon}_7(y)$ and $\Upsilon_4(x) + \Upsilon_8(x) = (x_3 + x_1)\mu'\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}\beta + x_2\frac{k^A_E}{V}\right) \geq (y_3 + y_1)\breve{\mu}'\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}\beta + y_2\frac{k^A_E}{V}\right) = \breve{\Upsilon}_4(y) + \breve{\Upsilon}_8(y)$. Let us now consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \; x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3, x_1 + x_4 = w_1 + w_4, x_1 + x_3 + x_4 \leq w_1 + w_3 + w_4\}$. Given that $\langle A_{3\bullet}, v_1\rangle = \langle A_{3\bullet}, v_9\rangle = 1$ and $\langle A_{3\bullet}, v_2\rangle = \langle A_{3\bullet}, v_{10}\rangle = 1$, we need to verify that $\Upsilon_1(x) + \Upsilon_9(x) \leq \breve{\Upsilon}_1(y) + \breve{\Upsilon}_9(y)$ and $\Upsilon_2(x) + \Upsilon_{10}(x) \geq \breve{\Upsilon}_2(y) + \breve{\Upsilon}_{10}(y)$ hold. Since $x_2 \geq y_2, x_1 + x_3 \leq y_1 + y_3, x_1 + x_4 = y_1 + y_4, x_1 + x_3 + x_4 \leq y_1 + y_3 + y_4$, then $\Upsilon_1(x) + \Upsilon_9(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_4))\left(k^2_{W0} + \frac{k_M}{V}(x_1 + x_4) + \frac{\bar{k}_M}{V}(x_1 + x_3)\right) \leq (\mathrm{D_{tot}} - (y_1 + y_2 + y_4))\left(k^2_{W0} + \frac{k_M}{V}(y_1 + y_4) + \frac{\bar{k}_M}{V}(y_1 + y_3)\right) = \breve{\Upsilon}_1(y) + \breve{\Upsilon}_9(y)$ and $\Upsilon_2(x) + \Upsilon_{10}(x) = (x_1 + x_4)\mu\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}b + x_2\frac{k^A_E}{V}\right) \geq (y_1 + y_4)\mu\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}b + y_2\frac{k^A_E}{V}\right) = \breve{\Upsilon}_2(y) + \breve{\Upsilon}_{10}(y)$. Finally, let us consider $x \in \mathcal{X}$ and $y \in \partial_3(K_A + x) \cap \mathcal{X} = \{w \in \mathcal{X} : \; x_2 \geq w_2, x_1 + x_3 \leq w_1 + w_3, x_1 + x_4 \leq w_1 + w_4, x_1 + x_3 + x_4 = w_1 + w_3 + w_4\}$. Given that $\langle A_{4\bullet}, v_7\rangle = \langle A_{4\bullet}, v_9\rangle = 1$ and $\langle A_{4\bullet}, v_8\rangle = \langle A_{4\bullet}, v_{10}\rangle = -1$, we need to verify that $\Upsilon_7(x) \leq \breve{\Upsilon}_7(y)$, $\Upsilon_9(x) \leq \breve{\Upsilon}_9(y)$, $\Upsilon_8(x) \geq \breve{\Upsilon}_8(y)$ and $\Upsilon_{10}(x) \geq \breve{\Upsilon}_{10}(y)$ hold. Since $x_2 \geq y_2, x_1 + x_3 \leq y_1 + y_3, x_1 + x_4 \leq y_1 + y_4, x_1 + x_3 + x_4 = y_1 + y_3 + y_4$, that also imply $x_1 \leq y_1, x_3 \geq y_3, x_4 \geq y_4$, then $\Upsilon_7(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3 + x_4))\left(k^1_{W0} + \frac{k'_M}{V}(x_1 + x_4)\right) \leq (\mathrm{D_{tot}} - (y_1 + y_2 + y_3 + y_4))\left(k^1_{W0} + \frac{k'_M}{V}(y_1 + y_4)\right) = \breve{\Upsilon}_7(y)$, $\Upsilon_9(x) = (\mathrm{D_{tot}} - (x_1 + x_2 + x_3 + x_4))\left(k^2_{W0} + \frac{k_M}{V}(x_1 + x_4) + \frac{\bar{k}_M}{V}(x_1 + x_3)\right) \leq (\mathrm{D_{tot}} - (y_1 + y_2 + y_3 + y_4))\left(k^2_{W0} + \frac{k_M}{V}(y_1 + y_4) + \frac{\bar{k}_M}{V}(y_1 + y_3)\right) = \breve{\Upsilon}_9(y)$, $\Upsilon_8(x) = x_3\mu'\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}\beta + x_2\frac{k^A_E}{V}\right) \geq y_3\breve{\mu}'\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}\beta + y_2\frac{k^A_E}{V}\right) = \breve{\Upsilon}_8(y)$ and $\Upsilon_{10}(x) = x_4\mu\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}b + x_2\frac{k^A_E}{V}\right)$
$\geq y_4\mu\left(\varepsilon \frac{k^A_M}{V}\mathrm{D_{tot}}b + y_2\frac{k^A_E}{V}\right) = \breve{\Upsilon}_{10}(y)$. Then, condition (ii) of Theorem S.2 in [5] also holds.

We can then conclude that all of the conditions of Theorem S.2 in [5] hold and so do the conclusions of Theorem 3.4 in [5], as per the remarks in SI - Section S.3 in [5].

REFERENCES

[1] Asmussen, S. (2003). *Applied Probability and Queues, 2nd edition.* Stochastic Modelling and Applied Probability, Springer, New York, NY.

[2] Avrachenkov K.E., Filar, J.A. and Howlett, P.G. (2013). *Analytic Perturbation Theory and Its Applications.* SIAM, Philadelphia.

[3] Bruno, S., Williams, R.J. and Del Vecchio, D. (2022). *Epigenetic cell memory: The gene's inner chromatin modification circuit.* PLOS Computational Biology, Public Library of Science, vol. 18(4): 1 - 27.

[4] Meyer, Jr. and Carl, D. (1975). The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, vol. 17, no. 3, pp. 443–464.

[5] Campos, F.A., Bruno, S., Fu, Y., Del Vecchio, D. and Williams, R.J. (2023). Comparison theorems for stochastic chemical reaction networks. *Bull Math Biol*, vol. 85(39).

[6] Friedberg, S. H., Insel, A. J. and Spence, L. E. (2014). *Linear Algebra, 4th edition.* Pearson Education.

[7] Gardiner, C.W. (1994). *Handbook of stochastic methods for physics, chemistry and the natural sciences.* Springer-Verlag.

[8] Gaver, D. P., Jacobs, P. A. and Latouche, G. (1984). Finite birth-and death models in randomly changing environments. *Advances in Applied Probability*, Vol. 16, 715-731.

[9] Grassmann, W.K. (1977). Transient solutions in Markovian queueing systems. *Computers & Operations Research*, vol. 4, no. 1, pp. 47–53.

[10] Horn, R.A and Johnson, C.R. (2013). *Matrix Analysis, 2nd Edition* Cambridge University Press.

[11] Issacson, D.L. and Madsen R.W. (1985). *Markov Chains: Theory and Applications* R.E. Krieger Pub. Co.

[12] James, M. (1978). The generalised inverse. *The Mathematical Gazette,* 62 (420), 109-114.

[13] Kelly, F.P. (1979). *Reversibility and Stochastic Networks.* Cambridge University Press.

[14] Kemeny, J.G. and Snell, J.L. (1960). *Finite Markov Chains.* Van Nostrand Comp. Int., New York

[15] Liggett, T.M. (2010). *Continuous Time Markov Processes: An Introduction.* Graduate Studies in Mathematics, Volume 113 . American Mathematical Society.

[16] Norris, J.R. (1997). *Markov Chains.* Cambridge University Press.

[17] Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3), 406-413.