Supplementary Material: Transcriptional delay stabilizes bistable gene networks

Chinmaya Gupta,[1] José Manuel López,[1] William Ott,[1] Krešimir Josić,[1,2] and Matthew R. Bennett[3,4,5]

[1] *Department of Mathematics, University of Houston, Houston, TX*
[2] *Department of Biology & Biochemistry, University of Houston, Houston, TX*
[3] *Department of Biochemistry & Cell Biology, Rice University, Houston, TX*
[4] *Institute of Biosciences & Bioengineering, Rice University, Houston, TX*
[5] *To whom correspondence should be addressed: matthew.bennett@rice.edu*

Outline. We present here the details of several computations that are described in the main manuscript. We also describe in detail all the models that have been used in the manuscript, and expand on various notes in the discussion in the manuscript.

First, we present the detailed analysis of the reduced model, and derive the primary expression of interest. Second, we outline the modifications to the stochastic simulation algorithm required to incorporate delay (page 3). Third, we provide a discussion of the positive feedback model, and present the parameters used for simulation (page 3), as well as discuss the changes in the stability of the fixed points for the deterministic approximation (page 3). We also discuss the case of distributed delays (page 4), and delayed deaths (page 4). Fourth, we present the details of the lysis/lysogeny switch of phage λ (page 5). Finally, we present the details of the co-repressive toggle switch, along with a discussion of a geometric method for varying the fixed points of the corresponding deterministic system in terms of the system parameters (page 7). The qualitative stabilization effect is, however, independent of the system parameters.

In all models time is scaled so that one unit of time corresponds to the half-life of a protein.

ANALYSIS OF THE REDUCED MODEL (RM)

We recall here the structure of the RM and derive the various estimates of interest. The states L and H in the RM correspond to the those regions of the phase space that are in the vicinity of a stable point. While the process is in the potential well, it undergoes small fluctuations around the fixed points of the corresponding deterministic system.

The third state, I, is an intermediary state. All transitions from the basin of one stable point must cross through the intermediary state before they can fall into the second basin. In the phase space, the intermediary state is represented by a thick neighborhood of the separatrix for the basins of attraction. As described in the main manuscript, since the system retains memory, the immediate history is an important consideration around the separatrix. A fluctuation may push the current state of the system from one basin to another, but mature molecules entering the population can push the system back into the old potential well. This is markedly different from the Markovian case, where only the current population composition is needed for determining future probabilities.

The idea behind the analysis of the RM is to first discretize time using a step size Δ and to obtain the probability of making failed transitions for a given delay τ. The probability density function for the random variable that counts the number of failed transitions in the continuous limit is obtained by taking the limit $\Delta \to 0$. This, in turn, allows us to compute the mean number of failed transitions and the mean time spent in a failed transition, as well as the mean time spent in a successful transition, assuming that the residence times dominate the delay. These are the primary ingredients required for the computation of the residence times in the stable states.

We study the discrete case first. If the delay is assumed to be $K\Delta$ (where K is some positive integer), the RM can be embedded in a $(K+1)$-dimensional space, and can be represented by vectors $\vec{x} = (x_{-K\Delta}, x_{-(K-1)\Delta}, \ldots, x_{-\Delta}, x_0)$ with the state x_0 being the current state of the RM and $x_{-t\Delta}$ being the state t steps in the past. In the higher-dimensional space, not all jumps are feasible, and the process can only possibly jump from a configuration $\vec{x}$ to a configuration $\vec{y}$ if $x_{-i\Delta} = y_{-(i+1)\Delta}$ for all $0 \le i \le K-1$. The delay can now be expressed by saying that the probability of a feasible jump from $\vec{x}$ to $\vec{y}$ is given by $\Lambda_{x_0 \to y_0}^{x_{-K\Delta}}$. Call f_H the probability of failing a transition, given that the transition starts in the state H, and $x_{-i\Delta} = H$ for all $0 \le i \le K$. Let $P(k)$ denote the probability mass function for the number of steps k that are required to complete the loop $H \to I \to H$ given that $H \to I$ has occurred. We see that

$$P(k) = \frac{1}{f_H} \begin{cases} (\Lambda_{I\to I}^H)^{k-1}(1-\Lambda_{I\to I}^H)\frac{\Lambda_{I\to H}^H}{\Lambda_{I\to H}^H+\Lambda_{I\to L}^H} & 1 \le k \le K; \\ (\Lambda_{I\to I}^H)^K(\Lambda_{I\to I}^I)^{k-K-1}(1-\Lambda_{I\to I}^I)\frac{\Lambda_{I\to H}^I}{\Lambda_{I\to H}^I+\Lambda_{I\to L}^I} & k \ge K+1. \end{cases} \tag{S1}$$

	H	I	L
$\tau = 0$	0.502	2.019×10^{-5}	0.498
$\tau = 0.04$	0.498	1.977×10^{-5}	0.502
$\tau = 0.08$	0.500	2.045×10^{-5}	0.499

TABLE I: Shown here are the stationary distributions for the RM for three different values of τ: $\tau = 0, \tau = 0.04$ and $\tau = 0.08$. Rates used in the RM are $\lambda^I_{I\to x} = 50, \lambda^x_{I\to x} = 99, \lambda^x_{I\to x^c} = 1, \lambda^I_{x\to I} = 0.20, \lambda^x_{x\to I} = 0.0002, \lambda^x_{x^c\to I} = 0.004; x \in \{H, L\}; x^c = H$ if $x = L$ and $x^c = L$ if $x = H$.

In the continuous-time limit, a discrete delay $K\Delta$ is replaced by a delay τ where $\Delta \to 0$ and $K\Delta \to \tau$. As in the discrete case, for the continuous-time system let f_H denote the probability of failing a transition, given that the transition initiates from state H and the process remembers only state H when the transition begins. Let $P(t)$ denote the probability density function for the random variable F_H: the time needed to complete the $H \to I \to H$ loop given that $H \to I$ has occurred.

For the continuous-time case, we compute a formal limit in Eq. (S1). To do so, we set up some notation. For a continuous-time Markov process, a transition probability in a time interval of length Δ is given by $\lambda\Delta$ were λ is the corresponding transition rate for the process. In the discrete-time description of the process, we can replace probabilities such as $\Lambda^j_{i\to k}$ $(i \neq k)$ by $\lambda^j_{i\to k}\Delta$ (rates corresponding to transitions) and probabilities $\Lambda^j_{I\to I}$ by $(1 - (\lambda^j_{I\to H} + \lambda^j_{I\to L})\Delta)$ for $j \in \{H, I, L\}$. For fixed t and any Δ such that $t\Delta^{-1}$ is a positive integer, Eq. (S1) gives

$$P(F_H \in [t-\Delta, t]) = \frac{1}{f_H}\begin{cases} \left(1 - \left(\lambda^H_{I\to H} + \lambda^H_{I\to L}\right)\Delta\right)^{t\Delta^{-1}-1} \lambda^H_{I\to H}\Delta & \Delta \le t \le \tau \\ \left(1 - \left(\lambda^H_{I\to H} + \lambda^H_{I\to L}\right)\Delta\right)^{\tau\Delta^{-1}} \left(1 - \left(\lambda^I_{I\to H} + \lambda^I_{I\to L}\right)\Delta\right)^{(t-\tau)\Delta^{-1}-1} \lambda^I_{I\to H}\Delta & t \ge \tau + \Delta. \end{cases} \tag{S2}$$

Taking $\Delta \to 0$, we obtain

$$P(t) = \frac{1}{f_H}\begin{cases} \lambda^H_{I\to H} \exp\left(-(\lambda^H_{I\to H} + \lambda^H_{I\to L})t\right) & 0 < t \le \tau \\ \lambda^I_{I\to H} \exp\left(-(\lambda^H_{I\to H} + \lambda^H_{I\to L})\tau - (\lambda^I_{I\to H} + \lambda^I_{I\to L})(t-\tau)\right) & t > \tau \end{cases}. \tag{S3}$$

Since $P(t)$ is a pdf, it follows from integrating on $[0, \infty)$ that

$$f_H = (1 - Z_H(\tau))p^H_{I\to H} \quad + \quad Z_H(\tau)p^I_{I\to H}$$

where

$$p^i_{I\to H} = \frac{\lambda^i_{I\to H}}{\lambda^i_{I\to L} + \lambda^i_{I\to H}}, \quad Z_H(\tau) := \exp(-(\lambda^H_{I\to H} + \lambda^H_{I\to L})\tau).$$

Analogous computations can be performed for the probability of failing a transition if it is assumed that the transition initiates from state L, and the only remembered state is L. Analogous computations can also be performed to obtain the probability density function for the time spent in making a successful transition loop $H \to I \to L$, and, therefore, the probability of a successful transition. Further, we can compute the expectations of F_H, and S_H (defined as the time spent in a successful transition):

$$E[F_H] = \int_0^\infty tP(t)\, dt$$

et cetera. The exact forms of these expressions can be computed analytically, but we do not write them here. Instead, we note that $dE[F_H]/d\tau$ and $dE[S_H]/d\tau$ are very small. This implies that the times spent in state I during a failed transition, or a successful one, do not significantly change with τ.

Denote by N the number of failed transitions before a successful transition. Then N has a geometric distribution $(P(N = n) = f^n_H(1 - f_H), n \ge 0)$.

In between each failed transition, the process spends time in the stable states H or L before jumping out on another excursion. We have assumed that the residence times sufficiently dominate the delay; a consequence of this assumption is that once the process re-enters the state H, it stays there long enough to forget its past excursions. Therefore, we can estimate the time between transition attempts by $1/\lambda$ where $\lambda = \lambda^H_{H\to I}$.

We can now estimate the expected residence time in the stable state:

$$E[R_H] \sim \frac{f_H}{1 - f_H}\left(E[F_H] + \frac{1}{\lambda}\right) + E[S_H] + \frac{1}{\lambda}.$$

GILLESPIE'S STOCHASTIC SIMULATION ALGORITHM WITH DELAY

Gillespie's stochastic simulation algorithm (SSA) is a way to sample exact stochastic realizations of a chemical system. In the SSA, the reactions are modeled as birth-death processes. In the classical SSA at time t the state of the system is described by the population vector $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$. The set of possible reactions in the system is indexed by $\{1, 2, \ldots, M\}$. For each reaction j, a propensity function $a_j : \mathbb{R}^N \to \mathbb{R}^+$ describes the rate at which the reaction fires, given the population $\mathbf{x}(t)$. A vector $\mathbf{v}_j \in \mathbb{Z}^N$ describes the change in each species when reaction j fires. The total rate of all reactions together is given by $\sum_j a_j(x(t))$. The SSA proceeds by first sampling a time, Δ, to the next reaction from the exponential distribution with mean $\sum_j a_j$. The reaction is assigned type i with probability $a_i / \sum_j a_j$. The population $\mathbf{x}(t)$ is then updated by adding the appropriate state change vector $\mathbf{x}(t) \mapsto \mathbf{x}(t) + \mathbf{v}_i$.

In the examples considered in the manuscript some reactions (in most of the manuscript these are births) only affect the population size after a delay, while others (in most of the manuscript these are deaths) affect the population size immediately. Delays need not be of constant length. In the case of non-constant delays we assume that they are i.i.d. with distribution $\kappa(\tau)$.

To simulate such processes we used a modified version of Gillespie's algorithm [1]: At time t in the simulation the time to the next reaction, Δ, and the type of reaction, i, is sampled, as described above. Before proceeding Δ units of time, one checks to see if there are any reactions that commenced in the past, and finish in the interval $[t, t + \Delta]$. If there are no such reactions, one proceeds to time $t + \Delta$. If the reaction sampled, i, is a non-delayed reaction, the population is updated immediately. If i is a delayed reaction, the state change vector corresponding to i is put in a queue, with a designated time of exit τ units of time in the future, sampled from some distribution $\kappa(\tau)$.

If on the other hand, there is a reaction from the past which terminates in the time $[t, t + \Delta]$, one proceeds to the time of this reaction. The population size is changed according to the state change vector of this past reaction. The original waiting time Δ, and reaction type i, are discarded, and a new waiting time and reaction type are sampled.

POSITIVE FEEDBACK MODEL

The deterministic delay-differential equation that approximates the stochastic dynamics of the single gene positive feedback model is given by

$$\dot{x} = \alpha + \beta \frac{x(t-\tau)^b}{c^b + x(t-\tau)^b} - \gamma x \tag{S4}$$

The parameters used were $\beta = 20$, $\alpha = 5$, $c = 19$, $\gamma = \ln(2)$ and $b = 10$. The parameter α is the basal rate of production of the molecules of x (with units molecules s^{-1}). The maximal rate of production for the system is $\alpha + \beta$. c is the number of molecules of x required to achieve the half-maximal activation rate of $\alpha + \beta/2$. The constant b is the Hill coefficient for the activation function, and γ is the rate constant for the degradation of the protein x (with units s^{-1}).

For this set of parameters, there are three fixed points for the delayed differential equation: $x = 7.2, x = 18.0$ and $x = 36.0$. The fixed points at 7.2 and 36.0 are stable, while the fixed point at $x = 18.0$ is unstable.

We map the phase space of the positive feedback model onto the RM using $H = [23, \infty)$, $L = [0, 13]$ and $I = (13, 23)$. A trajectory that starts in the state H, and makes an excursion into I is said to have a successful transition if it reaches state L before state H.

All trajectories are initialized in the state H at $x = 25$. The initial molecule production queue is assumed to be empty. A transient is computed for $(\tau^2 + 1) \times 10^4$ units of time. Any data is gathered after the transient. The mean residence times R_τ are computed by averaging over 10^4 transitions.

Stability analysis for the Positive Feedback Model. We analyze the spectrum of the linearization around the fixed point of the delay differential equation (S4) rewritten as

$$\dot{x}(t) = B(x(t-\tau)) - D(x). \tag{S5}$$

This equation is the deterministic counterpart of the stochastic positive feedback model examined in the manuscript. Linearizing Eq. (S5) in the neighborhoods of a stable fixed point x_0 to yield a DDE

$$\dot{x}(t) = B'(x_0)(x(t-\tau) - x_0) + D'(x_0)(x(t) - x_0).$$

On setting $y(t) = x(t) - x_0$, $p = -B'(x_0)$, $q = D'(x_0)$, and assuming a solution of the form

$$y(t) = Ce^{st}$$

we get a characteristic equation

$$(s + q)e^{s\tau} + p = 0.$$

The k^{th}pair of eigenvalues s_k can be obtained by solving the equation

$$s_k = \frac{1}{\tau} W_k(-p\tau e^{\tau q}) - q$$

where W_k is the k^{th} branch of the Lambert W function. If s_0 is found to have negative real part, no other eigenvalues need to be computed as the stability is determined completely by s_0.

As is apparent from Fig S1, as τ increases, the stability of both stable fixed points decreases. Therefore, while the stochastic positive feedback system becomes mores stable, its deterministic counterpart becomes less stable with an increase in delay.

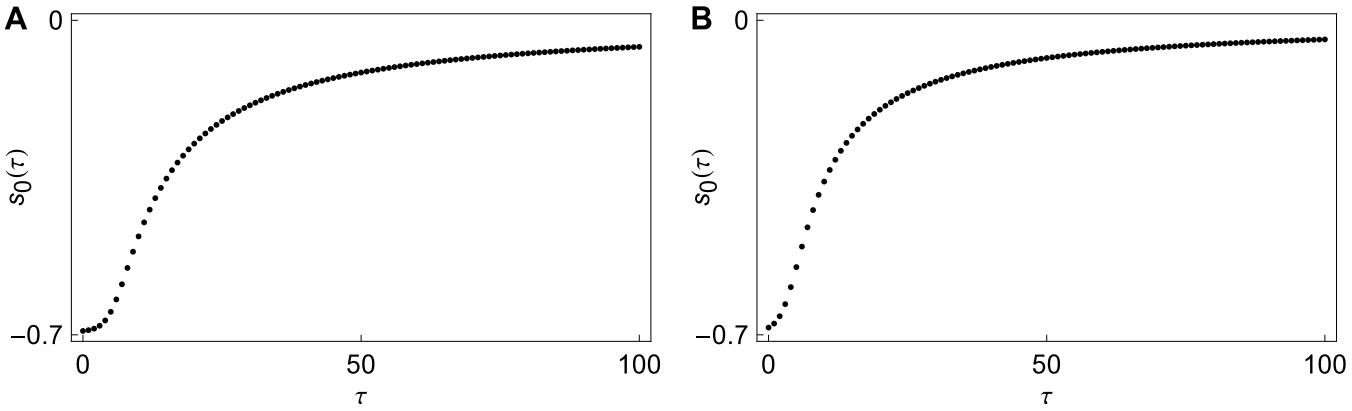

FIG. S1: **Left:** Plot of the leading eigenvalue, as a function of the delay, for the linearization of the system in the neighborhood of the lower fixed point $x_0 = 7.21$. **Right:** Plot of the leading eigenvalue, as a function of the delay, for the linearization of the system in the neighborhood of the higher fixed point $x_0 = 36.01$.

Distributed Delays. To examine the effect of distributed delay, we used Gamma delay distributions, $\kappa(\tau; \mu, \sigma)$, with different means and variances. The effect of increasing the mean of the distribution κ was qualitatively similar to increasing the magnitude of a constant delay, τ: The mean residence times increased sharply with small increases in this mean, and eventually appeared to saturate.

Increasing the variance of the gamma distribution for a fixed mean appears to initially slow down the rate of increase of the mean transition time with increasing delay; however, larger variances also appear to correspond to larger saturation values (see Fig. S2).

Delayed Deaths. We also consider a process that is formally constructed by delaying deaths (see Fig. S3). In such a model, as in the stochastic simulation algorithm with delayed births, the waiting time to the next reaction is sampled. If that reaction is a birth type reaction, the population is updated immediately. For a degradation, we put the corresponding state change vector in a queue with a designated time of exit. A new reaction time is then sampled. Otherwise the algorithm is as described above.

Delaying reactions that decrease population is less realistic, and can lead to negative population sizes. We disregard any reactions that would decrease the size of a population below 0. Formally, the deterministic equations approximating the stochastic dynamics in the case of delayed deaths can be written as

$$\dot{x} = \alpha + \beta \frac{x(t)^b}{c^b + x(t)^b} - \gamma x(t - \tau) \tag{S6}$$

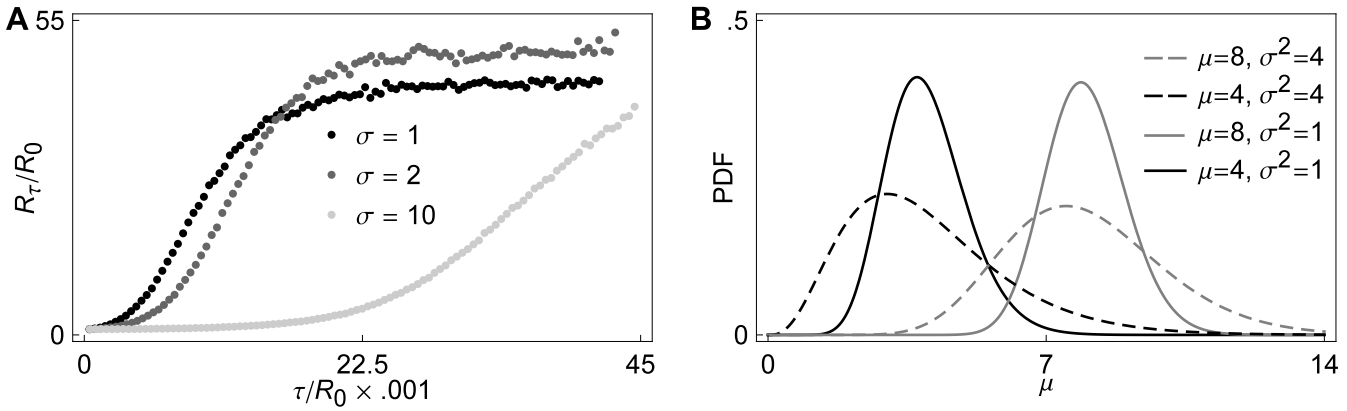

FIG. S2: **Left: Positive feedback model with distributed delay.** A Gamma distribution with $\tau \in (0, 10)$ and $\sigma \in \{1, 2, 10\}$ is used for the positive feedback model. The parameters used are $a = 20, b = 5, k = 19$. The trajectories are initialized with a population of size 25, followed by a long transient. For small values of τ, the ratio R_τ/R_0 grows more slowly for larger variances. **Right: Delay distributions.** We use families of Gamma distributions with varying means and variances $\sigma^2 = 1, 4$. Displayed here are the distributions for means 4 and 8.

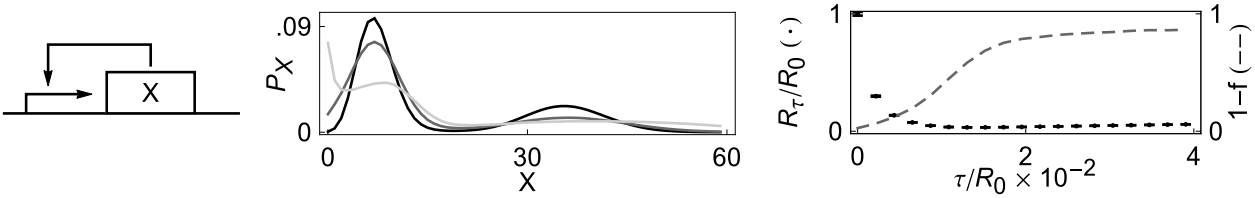

FIG. S3: **Effect of delaying degradation.** *(Left)* The motif for the positive feedback model. *(Center)* The stationary distributions for three different delays. Darker lines correspond to larger delays. The maximum accumulation of proteins of type X increases with delay. Since the exiting state change vectors depend on the protein numbers at a time in the past, there is an accumulation of probability at $X = 0$ with increasing delay. *(Right)* Solid dots represent the mean residence times. Dashed lines represent the probability of succeeding in an attempted transition.

with the added constraint that $x(t) \geq 0$ for all $t \geq 0$.

Delaying deaths destabilizes the bistable switch. The explanation parallels that given in the main manuscript:Consider a large downward fluctuations from away from the upper stable state H. In the presence of transcriptional delay, birth rates are determined by the state $x(t - \tau)$, and the larger τ, the more likely it is that $x(t - \tau)$ is in state H. But death rates in state H are high and favor motion away from H. Therefore, the trajectory is pushed away from the stable state it came from. The result is a large decrease in residence times with increasing delay.

The RM described in the manuscript can capture this effect by appropriately changing the transition rates $\lambda^H_{I\to H}$ etc.

THE LYSIS/LYSOGENY SWITCH OF PHAGE λ.

The lysis/lysogeny switch of phage-λ is realized as a set of chemical reactions involving multimerized forms of two transcription factors A, and B, which regulate gene expression by binding to the genome at an operator site O. The state of the operator is denoted by O if neither multimerized transcription factors are bound to it; when the multimer A_n is bound, the operator is denoted OA_n and when B_m is bound, the operator is denoted by OB_m.

A simplified model of the system consists of the following chemical reactions. The first two reactions represent the multimerization reactions of the transcription factors A and B. Multimerization is introduced into the model because transcription factors must bind to the DNA cooperatively to make a working switch.

$$nA \underset{k_b}{\overset{k_f}{\rightleftharpoons}} A_n \tag{S7a}$$

$$mB \underset{k_b}{\overset{k_f}{\rightleftharpoons}} B_m \tag{S7b}$$

The next two reactions represent the degradation of the transcription factor monomers as a first order reaction.

$$A \xrightarrow{\mu_A} \emptyset \tag{S8a}$$

$$B \xrightarrow{\mu_B} \emptyset \tag{S8b}$$

When no multimers are bound to the operator O, either A or B can be produced. This production is a result of a large number of biochemical steps, and for this reason, this reaction is assumed to be delayed. This delay is represented in the equations by setting τ as a superscript in the reaction rate constant. The synthesis reactions for A and B are then represented as

$$O \xrightarrow{k_A^{(\tau)}} O + A \tag{S9a}$$

$$O \xrightarrow{k_B^{(\tau)}} O + B. \tag{S9b}$$

The multimers A_n or B_m can bind reversibly to the operator O. This gives us

$$O + A_n \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} OA_n \tag{S10a}$$

$$O + B_m \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} OB_m. \tag{S10b}$$

Once the transcription factors of a certain kind bind to the operator, they allow only for the production of monomers of their own kind. This is represented by

$$OA_n \xrightarrow{k_A^{(\tau)}} OA_n + A \tag{S11a}$$

$$OB_m \xrightarrow{k_B^{(\tau)}} OB_m + B. \tag{S11b}$$

Together, this gives us a complete description of the lysis/lysogeny switch of phage$-\lambda$. We used the parameters $k_b = 5 = k_f = k_{on}, k_{off} = 1, k_A = 1 = k_B, \mu_A = 0.3 = \mu_B$ and $n = m = 2$. We assume the presence of only 1 operator O. A detailed analysis of the system, as well as a discussion of the model reduction of the deterministic approximation to a system of two ordinary differential equations can be found in [2].

The phase space of the lysis/lysogeny switch is mapped onto the RM as follows. We denote by N_X the number of molecules of type X in the system, we compute $|A| = N_A + 2N_{A_2}$ and $|B| = N_B + 2N_{B_2}$. If $|A| - |B| \geq 5$, the system is considered to be in state H, and if $|A| - |B| \leq 5$, in state L. Otherwise, if $|A| - |B| \in (-5, 5)$, the system is said to be in state I. The system is initialized with 5 molecules of the protein A and 30 molecules of A_2 (making $|A| = 65$) and $|B| = 0$. The initial molecule production queue is assumed to be empty. A transient is computed for $(\tau^2 + 1) \times 10^4$ units of time. Any data is gathered after the transient. Mean residence times are computed by averaging over 10^4 transitions.

CO-REPRESSIVE TOGGLE SWITCH.

We first describe the deterministic system in order to obtain the critical points. An appropriate choice of parameters is important, since transitions are very rare between stable points which are widely separated. We present here a geometrical method to easily find feasible parameters.

The standard form of the co-repressive toggle switch with delayed production is given by the set of delay ordinary differential equations [3]

$$\dot{x} = \frac{\beta}{1 + y(t-\tau)^2/k} - \gamma x$$
$$\dot{y} = \frac{\beta}{1 + x(t-\tau)^2/k} - \gamma y.$$

The production and degradation propensity functions for the delayed Gillespie algorithm are obtained from these ODEs. Since the fixed points of the system do not change if delay is introduced, we assume that $\tau = 0$, and parametrize the critical points of the non-delayed system: Setting $\dot{x} = \dot{y} = 0$ and writing $a = \beta/2\gamma$ we get

$$0 = \frac{\beta}{1 + y^2/k} - \gamma x = \frac{2ak}{k + y^2} - x$$
$$0 = \frac{\beta}{1 + x^2/k} - \gamma y = \frac{2ak}{k + x^2} - y.$$

We now eliminate y between the two equations to obtain

$$\frac{\left(2ak - kx - x^3\right)\left(k + x^2 - 2xa\right)}{\left(k + x^2\right)^2 + 4ka^2} = 0.$$

The numerator is a polynomial of degree 5, which implies that there exist at most 5 real critical points. Solving the quadratic part of the above equations, and substituting back into the original equation leads to two critical points if $a > \sqrt{k}$:

$$\left(a - \sqrt{a^2 - k}, a + \sqrt{a^2 - k}\right)$$
$$\left(a + \sqrt{a^2 - k}, a - \sqrt{a^2 - k}\right).$$

On solving the cubic term explicitly, we observe that we get only one real solution. We arrange for this solution to be on the line $y = x$ by choosing $k = s^3/(2a - s)$ (in which case the third critical point is (s, s)). On eliminating a, k from the above equations, we see that all critical points must lie on

$$xy(x + y - s) = s^3.$$

A stability analysis shows that for any s, the critical point at (s, s) is unstable, while the critical points $\left(a - \sqrt{a^2 - k}, a + \sqrt{a^2 - k}\right)$ and $\left(a + \sqrt{a^2 - k}, a - \sqrt{a^2 - k}\right)$ are stable.

Parameters can now be chosen by first choosing $s > 0$, then $a > s$ and finally $k = s^3/(2a - s)$. We fix the parameter $\gamma = \ln(2)$ because we assume our units of time to be in terms of the protein half-life. Finally, we can solve for the parameter $\beta = 2\gamma a$.

The parameters used in our study are $s = 10, \gamma = \ln(2)$ and $a = 15.8202$. The phase space of the co-repressive toggle is mapped onto the RM as follows: Denote by X and Y the number of molecules of each protein type in the system. The region between the $y-$ axis and the 45° line that passes half-way between the saddle and the stable point in the region $Y > X$ is mapped onto the state H. The corresponding region between the x-axis and the 45° degree line in the $X > Y$ region is mapped into state L. All trajectories are initialized at the saddle (s, s). The initial molecule production queue is assumed to be empty. A transient is computed for $(\tau^2 + 1) \times 10^4$ units of time. Any data is gathered after the transient. All numerical estimates for mean residence times, failed transition probabilities, and stationary distributions are computed for 10^4 transitions from state H to L (and back).

Transition trajectories. Delay also widens the distribution of paths that lead to failed transitions, as well as the distribution of those paths that correspond to successful transitions. The changes in the densities of the failed transition paths appear to be more sensitive to delay. Since the RM is not constructed using the specific features of any of the models, this effect cannot be explained using our reduction. We present in Figure-S4 the densities of the paths that correspond to failed, and successful transitions, for the co-repressive toggle switch.

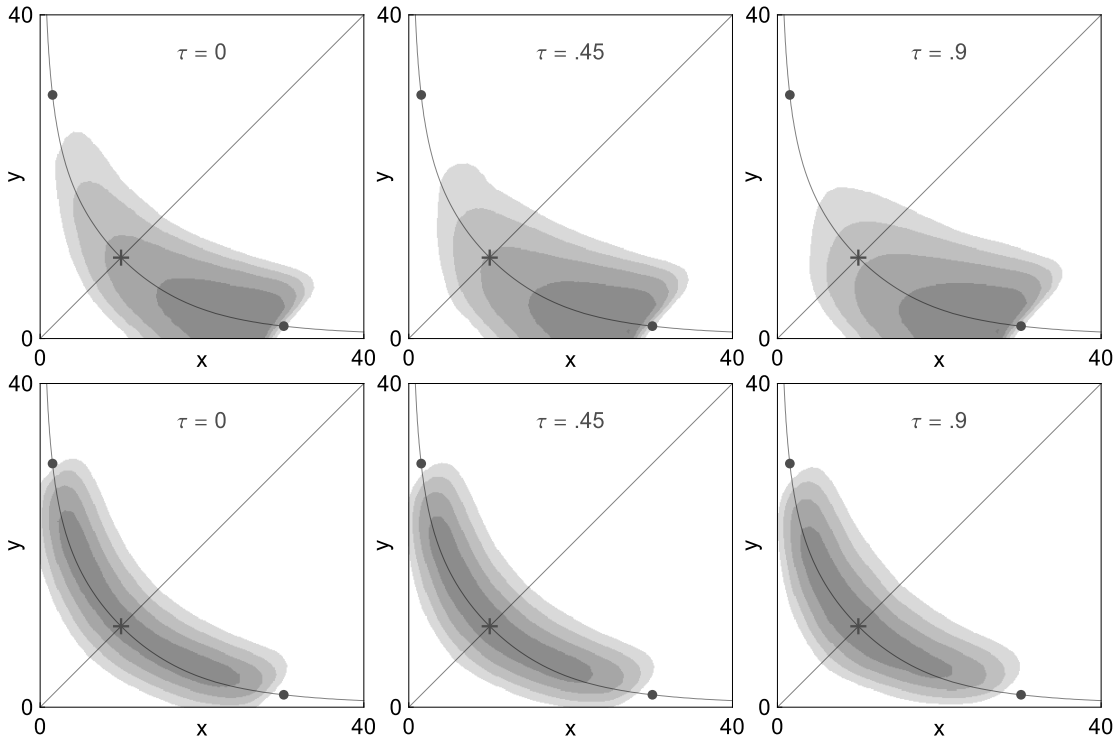

FIG. S4: The top panels show the density corresponding to trajectories that make a failed transition attempt, starting in a neighborhood of the stable point in the region $X > Y$. With increasing τ, the support of the density is more spread out. The bottom panels show the densities for the trajectories corresponding to successful transitions. Again, we observe that the support of the densities widens, although the effect is not as pronounced as in the case of failed transitions.

[1] R. Schlicht and G. Winkler, J Math Biol **57**, 613 (2008).
[2] P. Warren and P. Wolde, J Phys Chem B **109**, 6812 (2005).
[3] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature **403**, 339 (2000).