IOP Publishing Phys. Biol. 12 (2015) 045003 doi:10.1088/1478-3975/12/4/045003

Physical Biology

RECEIVED
5 November 2014

REVISED
26 February 2015

ACCEPTED FOR PUBLICATION
20 March 2015

PUBLISHED
18 June 2015

PAPER

Using noise to control heterogeneity of isogenic populations in homogenous environments

Paulina Szymańska[1,5], Nicola Gritti[2,5], Johannes M Keegstra[2,5], Mohammad Soltani[3,5] and Brian Munsky[4]

[1] Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawińskiego 5 b, 02–106 Warsaw, Poland
[2] FOM Institute AMOLF, Science Park 104, 1098XG, Amsterdam, The Netherlands
[3] Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA
[4] Department of Chemical and Biological Engineering and The School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, USA
[5] Contributed equally

E-mail: munsky@engr.colostate.edu

Keywords: stochastic gene expression, chemical master equation, toggle switch, single-cell control

Abstract

We explore the extent to which the phenotypes of individual, genetically identical cells can be controlled independently from each other using only a single homogeneous environmental input. We show that such control is theoretically impossible if restricted to a deterministic setting, but it can be achieved readily if one exploits heterogeneities introduced at the single-cell level due to stochastic fluctuations in gene regulation. Using stochastic analyses of a bistable genetic toggle switch, we develop a control strategy that maximizes the chances that a chosen cell will express one phenotype, while the rest express another. The control mechanism uses UV radiation to enhance identically protein degradation in all cells. Control of individual cells is made possible only by monitoring stochastic protein fluctuations and applying UV control at favorable times and levels. For two identical cells, our stochastic control law can drive protein expression of a chosen cell above its neighbor with a better than 99% success rate. In a population of 30 identical cells, we can drive a given cell to remain consistently within the top 20%. Although cellular noise typically impairs predictability for biological responses, our results show that it can also simultaneously improve controllability for those same responses.

1. Introduction

The overall goal for much of the biomedical sciences is to control biological behaviors to achieve a desired outcome (e.g., apply or maintain a sufficient dosage of chemicals to eliminate abnormal, cancerous cells) while minimizing deleterious effects (e.g., reduce toxicity, expense and inconvenience to the patient). In an ideal world, one would envision open-loop control laws, in which a fixed treatment regimen always achieves the same desired result without the need for intermediate changes or decisions. More realistically, every tissue, organism or person is different, and we will need to continually monitor and update our control strategies, giving rise to closed-loop control approaches where treatments are continually adjusted in time (e.g., personalized medicine). Although control of the immense complexity of biochemical reactions in multi-cellular organisms is beyond current imaginable technology, recent studies have sought to gain some insight into these processes through the analysis of simplified gene regulatory pathways in single-cells within larger populations [1, 2]. But even to control such simplified processes, one must first develop quantitative understanding for how control inputs interplay with the uncertainty associated with un-modeled biological complexities, and how these interactions affect the resulting biological responses.

One of the main impediments to development of precise and effective biomedical control strategies is the existence of uncertainties and variability. Issues of heterogeneities (commonly referred to as *noise*) often arise even at the level of single, genetically identical cells, which can exhibit diverse responses when exposed to identical environmental conditions or drug regimens [3–9]. This phenotypic variation is due in

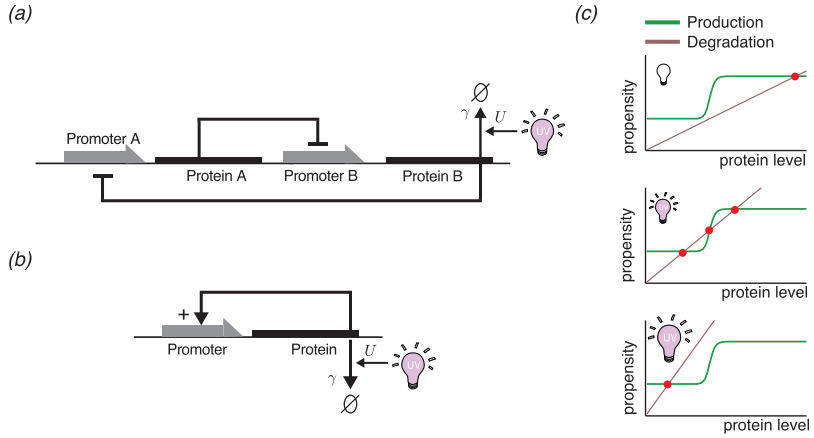

Figure 1. a) A classical representation of a genetic toggle switch [18], in which two protein species repress one another and the degradation rate of one protein is enhanced by UV radiation. (b) A simplified toggle switch model in which the two mutually inhibiting protein dynamics are approximated by a single, self-inducing gene. Protein degradation can be enhanced by increasing levels of UV radiation. (c) Production and degradation rates versus the number of proteins, x, in the simplified toggle switch model. The production rate, $k(x) = k_0 + k1^*x^m/(x^m + \beta^m)$, is plotted in green, and the degradation rate, $R(x) = (\gamma + U)x$, is plotted in brown. In a deterministic representation, protein levels evolve according to $\dot{x} = k(x) - R(x)$. Intersections where $k(x) = R(x)$ provide the stationary points. The three panels correspond to cases where UV is low (top, a single high equilibrium point), moderate (middle, two stable and one unstable equilibrium points) or high (bottom, a single low equilibrium point).

large part to the intrinsic stochasticity of gene expression and associated biochemical reactions within individual cells, and it has been shown that heterogeneity can have tremendous consequences on levels ranging well beyond single cells and up to the entire population [10–15]. Different phenotypes may be optimal or sub-optimal depending upon the specific environmental condition at each time. In environments subject to quick and potentially deadly changes, cells may exploit intrinsic variability and switch prematurely into phenotypes that are suboptimal for the current environment, but which are more likely to survive catastrophic environmental changes [5, 11, 16, 17]. Such bet-hedging strategies enable persistent bacteria to survive severe antibiotic treatments, or persistent cancer cells to resist chemotherapy [11]. By exploring the potential to selectively control the phenotypes of individual cells within a larger population of genetically identical cells, we hope to gain new insights on population behaviors. Such insight could be instrumental in the design of more effective control strategies in emerging biomedical and bioengineering applications.

In this work, we build simplified mathematical and computational models to capture the dynamics of stochastic gene regulatory responses when subjected to temporally fluctuating, yet spatially homogenous, environmental conditions. We then use this model to explore the theoretical possibility that one could select a single cell at random from within a population, and then use a single spatially uniform input to control that particular cell so that it achieves a different phenotype than the rest.

The particular gene regulatory network we study in this work is that of the synthetic genetic toggle switch, which was engineered by Gardner *et al* [18]. This system consists of two proteins, LacI and λcI, acting in a mutually inhibitory fashion (see figure 1(a)). For certain parameters sets, this design results in bistability, wherein each cell can be in one of two distinct states, where LacI is high and λcI is low, or where LacI is low and λcI is high. The LacI level of individual cells can be determined by means of a fluorescent protein reporter, which is expressed in tandem with LacI and is observed using fluorescence microscopy. In the work of Kobayashi *et al* [19], degradation of one of the regulatory proteins (λcI) was placed under the control of UV radiation via the SOS signaling pathway, which allowed the system to be pushed from the λcI-stable phenotype into the LacI-dominant phenotype. They found that in the absence of UV, cells remained in their original phenotypes; upon application of large quantities of UV, all cells switched to the low λcI state; and under short pulses of intermediate UV levels, some cells switched while others did not. This variability arises from randomness in the timing and order of cellular reactions, causing independent isogenic cells to exhibit slightly different levels of the λcI and LacI proteins at any point in time. Several stochastic models of the genetic toggle switch have been considered in the literature [20–22].

In recent studies, several groups have developed microfluidic based, real-time feedback control strategies aimed at using UV or chemical signals to control cellular behaviors over time [1, 2, 23–25]. These efforts have been aimed at controlling the entire population of cells or a single individual cell over time. Inspired by the genetic toggle switch [15, 17–19, 26] and the expanding ability to change environmental conditions based upon single-cell observations, we now explore the possibility to simultaneously drive the

bulk of a cell population to one phenotype while controlling a single individual cell to another.

In the following work, we first introduce a simplified mathematical description of the genetic toggle switch, which considers a single self-inducing protein. We then propose a control law of UV radiation, which applies to all cells equally, but in which the UV strength is chosen based on the relative states of the chosen cell and its surrounding neighbors. This control law introduces a coupling between the different cell trajectories. We show that a deterministic analysis predicts that separate control of individual cells is not possible. Next, using stochastic simulations and direct solutions of the chemical master equation, we show that the combination of cellular noise and an automatic control law can force a single chosen cell to maintain a qualitatively different response than the others. We illustrate this capability numerically first for a set of two cells and later for larger populations. To take into account protein production, folding and degradation, we examine the effects of time delays upon the results of the control law. We also explore the effects that measurement noise and extrinsic parameter variations can have on the successful control of a single cell in a population. Finally, we show that our approach designed using the simplified toggle switch can be extended to an expanded model of the genetic toggle switch including nonlinear interactions between two protein species within each cell.

2. Models and methods

In this work we consider a population of identical cells all subjected to the same environmental conditions. We start by simplifying the classical toggle switch mechanism [18], depicted schematically in figure 1(a), to a simpler model in which only one type of protein is produced and up-regulates itself (see figure 1(b)). In each cell, the protein of interest is produced at a basal rate, k_0, plus an induced rate via positive feedback, modeled by a Hill function of the form: $k_1 x^m/(x^m + \beta^m)$, where x is the instantaneous concentration of proteins, $k_0 + k_1$ is the saturated production rate and β is the concentration of protein at half maximum induction. Proteins are subject to natural degradation, proportional to their concentration. The differential equation for the time evolution of x in each cell thus reads:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k(x) - R(x) = k_0 + k_1\frac{x^m}{x^m + \beta^m} - \gamma x. \quad (1)$$

Provided that $m \geqslant 2$, it is possible for the ODE description of this system to realize either one (stable) or three (two stable and one unstable) stationary points. The actual number of equilibrium points depends upon the other parameters, and in particular the degradation rate, γ, a fact we will utilize below.

In their study of the genetic toggle switch, Kobayashi *et al* [19] constructed a synthetic circuit where the stability of the toggle state with high λcI could be disrupted through the application of UV radiation, which increased degradation of λcI. Along this line, we now assume that the system can be subjected to UV radiation, which enhances the instantaneous degradation rate of protein in our model according to $\gamma = \gamma_0 + u$. Under this control law, the differential equation for the protein concentration becomes:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_0 + k_1\frac{x^m}{x^m + \beta^m} - \left(u + \gamma_0\right)x. \quad (2)$$

We assume that all other parameters controlling protein production (m, k_0, k_1, β) and basal degradation (γ_0) are fixed and independent of UV. In this expression, the UV strength is the controllable input function, which depends solely on the protein number (concentration) and not explicitly on time. The UV signal is assumed to be continuous and differentiable, which guarantees the existence and uniqueness of the solutions for equation (2). Under these assumptions, u will determine the number and location of all stationary points, as given by the zeroes of the polynomial:

$$k_0 + k_1\frac{x^m}{x^m + \beta^m} - \left(u + \gamma_0\right)x.$$

Figure 1(c) shows three possible examples to compare the production rate, $k(x)$, and the degradation rate $R(x)$, where UV radiation is low (top), moderate (middle) or high (bottom). For low and high UV, there is a single equilibrium point at high or low values of x, respectively. For moderate UV, there are two stable and one unstable equilibrium points.

Now that we have a model for a single cell that can be switched via UV control from a single stable high equilibrium, to a pair of high and low stable equilibria, and then to a single low equilibrium, we now examine the effects of UV on multiple cells simultaneously. For this, we first consider a population with two independent cells, given by a two dimensional vector $\mathbf{x} = [x_1, x_2]$, where both x_1 and x_2 evolve according to the same ODE as before (equation (2)). Once again the number and location of equilibria depend upon the level of UV radiation. For low or high UV signals, x_1 and x_2 will converge to an equilibrium that is high or low, respectively. For intermediate UV levels, a total of four stable equilibria will be possible: low protein expression in both cells, high protein expression in both cells, low protein in cell one and high protein expression in cell two, or vice versa. Figures 2 (b)–(d) illustrate schematically how low, moderate or high UV radiations affect the phase diagrams of the protein trajectories for the two cells. In this deterministic setting, every control law, in which the UV level is a continuous and differentiable function of x_1 and x_2, will assign every initial condition to a unique corresponding trajectory. In particular, no trajectory can cross the $x_1 = x_2$ line. As a result, the ranking of gene expression

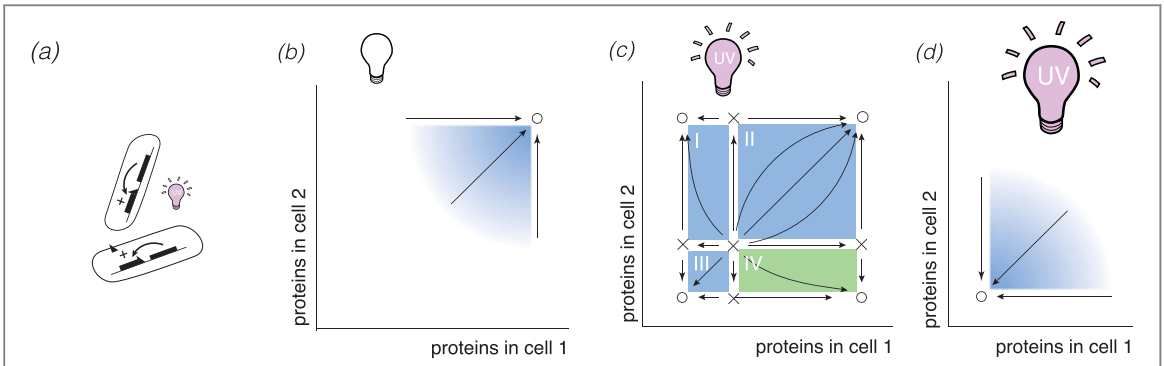

Figure 2. (a) A cartoon of the model considered: two identical cells each governed by a toggle switch are subjected to the same UV radiation. (b-d) A schematic phase diagram of the system with two cells, each governed by the same simplified toggle switch (see figure 1(b)) at three different UV radiation values: (b) low UV; (c) moderate UV; and (d) high UV. Axes represent number of proteins in cell one and cell two, circles and crosses mark the stable and unstable steady states, respectively.

levels in different cells depends exclusively on their initial condition, and it would be impossible to modulate this ranking using only a single control input. We next explore if the same conclusion holds when we account for intrinsic stochasticity in the underlying biochemical reactions.

2.1. Stochastic description

In our analysis, we assume that at any given time, all cells have identical parameters, and the UV radiation is applied equally to all cells. As a result, in our ODE description, if two genetically identical cells start at equal protein concentrations, then both will maintain equal protein concentrations for all subsequent times, and it is not possible to favor one cell over any of the others without direct intervention at the single-cell level. However, for systems involving small numbers of interacting molecules, cells may exhibit stochastic dynamics that are not captured by deterministic analyses. Instead, one must transition to a stochastic representation of protein dynamics, where protein numbers continually fluctuate over time and from one cell to the next. By random chance, a given cell will occasionally have higher protein expression than its immediate neighbor, and this could allow us to systematically alter the UV strength in order to maximize or minimize this difference.

In our stochastic description, the instantaneous number of proteins in N cells is described as a random non-negative integer vector, $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_N]^T \in \mathbb{N}^N$. Reactions change exactly one molecular species by a single molecule (positive or negative), and correspond to transitions from one state $\tilde{\mathbf{x}}_i$ to another $\tilde{\mathbf{x}}_j = \tilde{\mathbf{x}}_i \pm \mathbf{e}_h$, where the stoichiometry vector, $\mathbf{e}_h$, is zero except for the h^{th} entry, where it is unity. The propensities of the stochastic transitions (or stochastic reaction rates) are chosen using the same parameters and functions as for the deterministic process. Given the current state $\tilde{\mathbf{x}}$, the time until the next production or degradation event for species j is given by an exponential distribution with average waiting time $\mathbb{E}\{\tau_{+h}\} = k(\tilde{x}_h)^{-1}$ or $\mathbb{E}\{\tau_{-h}\} = (\gamma(u)\tilde{x}_h)^{-1}$, respectively. With these definitions of the stoichiometry and stochastic reaction rates, it is now possible to use the Stochastic Simulation algorithm (SSA, [27]) to simulate trajectories of the entire process for all N cells, and for any definition of the control law $u = u(\tilde{\mathbf{x}})$.

In these SSA simulations, each time a reaction fires, new stochastic reaction rates, $k(x_i)$ and $R(x_i, u)$, are computed for each cell, and the UV level is based on the instantaneous number of proteins in each cell, $u(x_1, x_2, \ldots)$.

As an alternative to the simulation of multiple trajectories, we can develop a chemical master equation to describe the probabilities of every possible state $\mathbf{x}_i = [x_{1i}, \ldots, x_{Ni}]^T$ in the infinite set of possible states $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$. Here, we use x_{hi} to denote the number of proteins in the h^{th} cell for the i^{th} state. Let $P_i(t)$ denote the probability that the system is in the i^{th} state at the time t, and let $\mathbf{P}(t) = [P_1(t), P_2(t), \ldots]^T$ be the vector of probabilities for all states $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$. This probability vector evolves according to the linear ordinary differential equation, $\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t)$, known as the chemical master equation, (CME). The infinitesimal generator, $\mathbf{A}$, in the CME is given by:

$$A_{ij} = \begin{cases} k_0 + k_1 x_{hj}^m \Big/ \left(\beta^m + x_{hj}^m\right) \\ \quad \text{for } \mathbf{x}_i = \mathbf{x}_j + \mathbf{e}_h \\ \left(\gamma + u(\mathbf{x}_j)\right) x_{hj} \\ \quad \text{for } \mathbf{x}_i = \mathbf{x}_j - \mathbf{e}_h \\ -\sum_{h=1}^{N} \left(k_0 + k_1 x_{hi}^m \Big/ \left(\beta^m + x_{hi}^m\right) \right. \\ \quad \left. + \left(\gamma + u(\mathbf{x}_i)\right) x_{hi} \right) \\ \quad \text{for } i = j \\ 0 \\ \quad \text{otherwise} \end{cases} \tag{3}$$

The elements of the infinitesimal generator matrix $\mathbf{A}$ include the dependence of $u(\mathbf{x})$ in the degradation rate at each point $\mathbf{x}_i \in \mathbf{X}$. We note that even though trajectories of $\tilde{\mathbf{x}}$ and $u(\tilde{\mathbf{x}})$ are discrete and random, the

CME is a linear time invariant ODE: $\dot{\mathbf{P}} = \mathbf{AP}$. Because the potential number of proteins in each cell can be any integer number, the dimension of the CME is infinite. However, we can approximate the solution by setting a limit on the maximum number of proteins allowed in each cell. By using an absorbing boundary condition at this limit, we can estimate the error incurred through this truncation at any time using the finite state projection approach [28]. Furthermore, by using a reflective boundary condition at this limit, we estimate the stationary distribution of the process as the solution of the linear equation $\mathbf{AP}_\infty = \mathbf{0}$. To integrate the truncated analysis of the CME forward in time, we use the package expokit [29, 30] implemented in Matlab [31]. For additional details on the set up and solution of discrete stochastic processes using the FSP approach, we refer the interested reader to references [28, 30, 32].

2.2. Two protein species model

We next consider an extension of the simplified toggle switch model to describe expression of two mutually repressive proteins in each cell. This represents the original toggle switch designed by Gardner *et al*, where the mutually inhibiting proteins are LacI and λcI [18]. A schematic for the standard toggle switch model is shown in figure 1(a). As redesigned by Kobayashi *et al* [19], the degradation rate of λcI is controlled via UV radiation, such that one can optogenetically push the system from a state of high λcI expression to a state of high LacI expression. We will later use this feature to specify a UV radiation control law such that the expression of LacI is made high in one pre-specified cell and small in all of the others. Using parameters originally fit to the original UV dependent response distributions [19, 20], the rates for the production of λcI and LacI are given by:

$$\begin{aligned} k_{\lambda\mathrm{cI}} &= k^{(0)}_{\lambda\mathrm{cI}} + \frac{k^{(1)}_{\lambda\mathrm{cI}}}{1 + \alpha_{\mathrm{LacI}}[\mathrm{LacI}]^{2.1}} \\ &= \left(6.8 \times 10^{-5} + \frac{0.016}{1 + 6.1 \times 10^{-3}[\mathrm{LacI}]^3}\right) s^{-1} \\ k_{\mathrm{LacI}} &= k^{(0)}_{\mathrm{LacI}} + \frac{k^{(1)}_{\mathrm{LacI}}}{1 + \alpha_{\lambda\mathrm{cI}}[\lambda\mathrm{cI}]^3} \\ &= \left(2.2 \times 10^{-3} + \frac{0.017}{1 + 2.6 \times 10^{-3}[\lambda\mathrm{cI}]^{2.1}}\right) s^{-1}. \quad (4) \end{aligned}$$

The degradation rate for LacI is the dilution rate due to division ($\gamma_{\mathrm{LacI}} = 3.8 \times 10^{-4} s^{-1}$), and the degradation rate for λ_{cI} depends upon the UV levels as:

$$\gamma_{\lambda\mathrm{cI}} = \begin{cases} 3.8 \times 10^{-4} s^{-1} & \text{at } UV = 0 \text{ J/m}^2 \\ 6.8 \times 10^{-4} s^{-1} & \text{at } UV = 6 \text{ J/m}^2 \\ 2 \times 10^{-3} s^{-1} & \text{at } UV = 12 \text{ J/m}^2 \end{cases} \quad (5)$$

For the stochastic simulation of the process in N genetically identical cells, the reactions are similar to those described above for the toy model, but the expanded state vector is now given by $\tilde{\mathbf{x}} = [\tilde{x}_{11}, \tilde{x}_{12}, \ldots, \tilde{x}_{N1}, \tilde{x}_{N2}]^T \in \mathbb{N}^{2N}$.

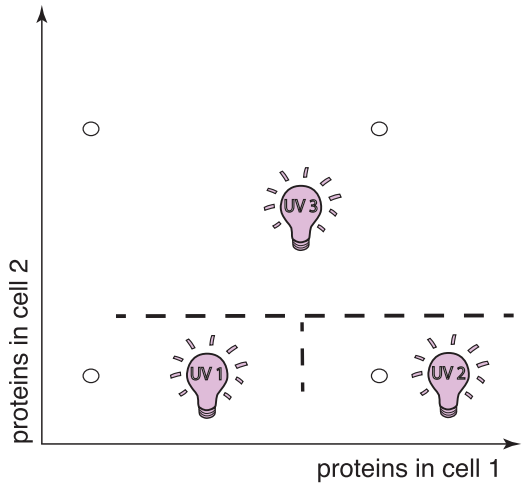

Figure 3. Preliminary control law for the application of UV radiation based upon comparison of the protein content for two cells. When both cells contain little protein (bottom left corner), no UV is applied ($u_1 = 0$). When protein content is high in cell one and low in cell two (bottom right corner) moderate UV is applied ($u_2 = 0.75$). When protein content in cell two is above a threshold (= 12) (upper part), high UV is applied ($u_3 = 1.75$). All parameters for the gene regulatory circuit ($k_0 = 5$, $k_1 = 50$, $\beta = 20$, $m = 4$ and $\gamma = 0.5$) are fixed and identical for both cells. Degradation rate parameters (i.e., γ and u_i) have arbitrary units of inverse time, and production rates (i.e., k_0 and k_1) have units of molecules per unit time.

3. Results

Beginning with the one-protein version of the toggle model and using our direct CME analysis, we found a simple control law for UV radiation that favorably drives a single chosen cell to express high levels of protein, while another identical cell is driven simultaneously to express low levels of protein. Figure 3 depicts the chosen strategy for how the UV level depends upon the level of protein in the chosen cell (x_1) and in the competitor (x_2). In this strategy, depending upon the pair (x_1, x_2), the UV signal at any given time attains one of three values: u_1, u_2 or u_3. This preliminary control law was built as follows: first, if the number of proteins in both cells is low (bottom left part of the plane), low UV is applied, allowing the cells to produce protein freely. Because of the symmetry between the cells, the trajectories will move roughly parallel to the $x_1 = x_2$ line. However, because of the stochastic nature of the process, the system will fluctuate into the regions where the protein expression is higher in one cell than the other. Second, if the cell population randomly moves to the region where protein expression in cell one is high, and the protein expression in cell two is low (bottom right part of the plane), we apply moderate UV to stabilize the system in this favorable condition. Third, when expression in cell two becomes elevated, we apply a higher level of UV. This reduces expression in both cells, but briefly

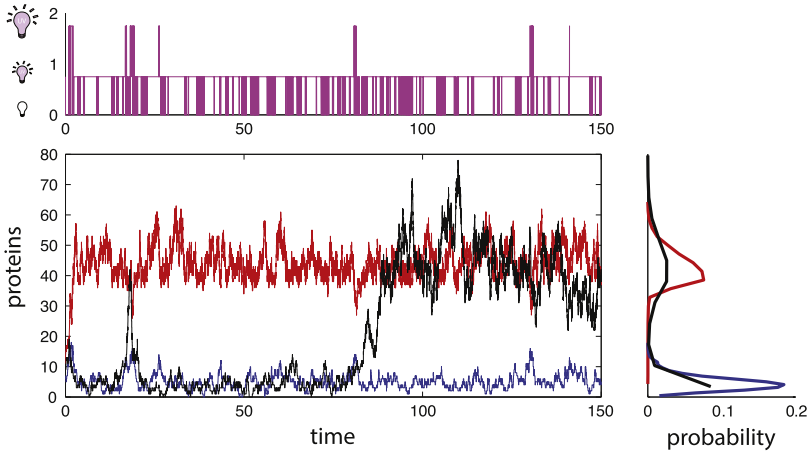

Figure 4. Example stochastic trajectories [27] for the number of proteins versus time for three cells, all subjected to the same UV signal over time. Trajectories for cells whose expressions are controlled to be high and low are shown in red and blue, respectively. A trajectory for a third cell that experiences the same UV signal but is not included in the control decisions, is shown in black. The top plot represents the UV radiation applied to all cells over time. The curves to the right show the marginal probabilities for the number of proteins in each cell, as determined form a single long stochastic trajectory.

and only until cell two falls below the threshold, thereby allowing cell one a greater chance to remain closer to the high stable state at moderate UV levels.

In figure 4, we illustrate the results of this control law when applied to the trajectories of proteins versus time for a population of three cells. These cells correspond to one cell intentionally driven to have high expression (red), one cell intentionally driven to have low expression (blue), and a third cell (black) that is subject to the fluctuating UV level but upon which the control law does not depend. Marginal distributions computed using a long time simulated trajectory are shown for each cell to the right of the trajectories. From the figure, it is clear that the UV control law successfully maintains the first cell in the high state (red, $\langle x_1 \rangle = 44.3 \pm 5.8$) and the second cell in the low state (blue, $\langle x_2 \rangle = 4.7 \pm 2.5$). Meanwhile, the cell that is not specifically controlled (black, $\langle x_3 \rangle = 23 \pm 21$) is free to fluctuate between high and low states.

In addition to simulating the trajectories for the two cells, we also used the direct chemical master equation analysis to integrate the CME forward in time and estimate the stationary marginal and joint distributions of the protein content of the two competing cells. Figure 5 summarizes these distribution shapes for four different cases of UV control: low, moderate, high, and using the control law defined in figure 3. In each case the joint distribution is shown on the left and the marginal distributions are shown on the right. UV levels that do not depend upon the concentrations in the two cells, shown in figures 5(a)–(c), do not break the symmetry between the cells, and the UV-dependent marginal distributions are identical for the two cells. In contrast, figure 5(d) shows that the chosen UV control law is successful to make it highly probable that cell one has high expression and cell two has low expression. Solving numerically the CME, we found that the equilibrium probability that the protein level in cell one exceeds that of cell two is $\mathrm{Prob}(x_1 > x_2) = 0.998$ with the specified control law. To confirm the consistency between the stochastic simulations and the direct CME solutions, figure 5(d) plots the marginal solutions for both approaches.

We next extended the control law found for the two-cell population to the case of a N-cell population, $N \geqslant 2$. To keep the control law as simple as possible, we maintained the same three regions of different UV values shown in figure 3, as well as the same three values for the UV radiation $\{u_1, u_2, u_3\}$. However, the UV level is now based on the comparison of the number of proteins in cell one, x_1, and the maximum number of proteins in all other cells, $\max_{i=2}^{N}(x_i)$. As above, we simulated the trajectories of all N cells simultaneously under the controlled UV radiation. Foreseeably, the effectiveness of the control law is decreased when a larger population is considered. This effect is quantified in figure 6(a), (top, black diamonds), which shows the time averaged percentile rank of the chosen cell versus the total number of cells in the population. The probability that protein expression in cell one is higher than the mean expression in all other cells drops below 80% only for $N \geqslant 30$.

In order to make the model more applicable to practical implementation, we next extended the analysis to consider time delays that are present in all cellular processes [33, 34], especially in the production and maturation of the fluorescent proteins that are frequently used to quantify gene regulatory responses [35]. To introduce the effects that such a time delay would have on the differential control of single cells, we assume that the UV control law depends not upon the state at the current time, $\mathbf{x}(t)$, but upon the state at an earlier time $\mathbf{x}(t-\tau)$. Figures 6(a) and 6(b) show how the effectiveness of the control law decreases as

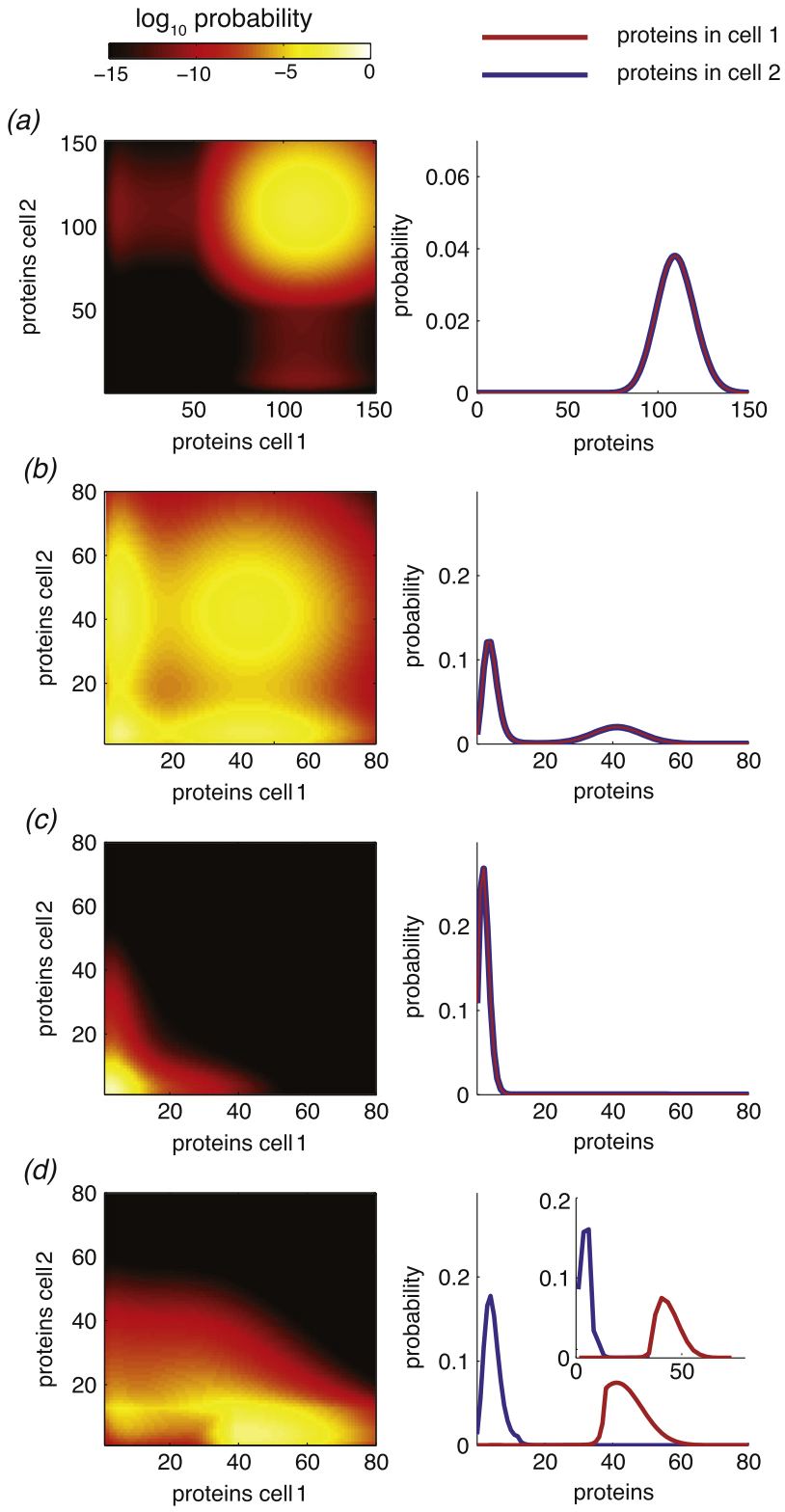

Figure 5. Stationary distributions for the joint probabilities (left) and marginal probabilities (right) for the protein content in two cells under four different UV control laws: (a) low UV, (b) moderate UV, (c) high UV and (d) UV control law given in figure 3. On the right, marginal distributions computed using CME, for cell 1 (red) and cell 2 (blue), are compared. The inset in panel (d) shows the marginal probability distributions computed from a single long trajectory obtained using the SSA [27].

the time delay increases (e.g., due to longer fluorescent maturation times or delays in the activation of the UV-induced SOS pathway). While short delays may be tolerated, once time delays reach or exceed the half-life of the controlled protein, the effectiveness of the control law rapidly diminishes. Maturation times for fluorescent proteins vary considerably from about two minutes for the fastest variant of yellow fluorescent protein to thirty or more minutes for common fast folding green fluorescent proteins [36, 37]. For

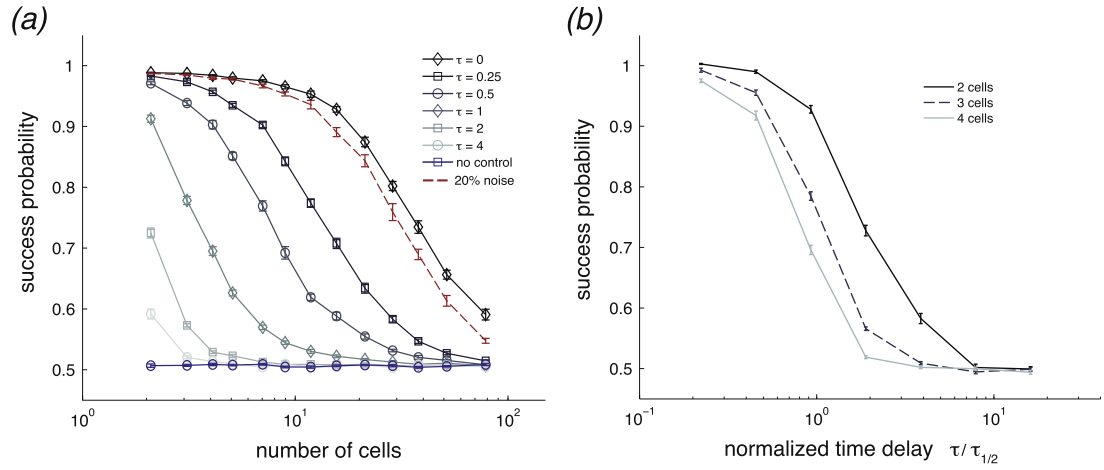

Figure 6. Time-averaged percentile rank of the chosen cell versus the number of competing cells in the population for measurement noise and time delays. The UV is chosen as discussed in the main text. (a) Success rates versus the number of cells in the population. Solid lines correspond to different measurement time delays, τ, using the same time units as figure 3. The dashed line corresponds to no time delay but with 20% standard deviation uncorrelated white noise added to all measurements. The blue line corresponds to the success rate in the absence of any control law. (b) Success rates versus the time delay normalized by the protein half-life at the lowest UV setting $\tau/\tau_{1/2} = \tau/(\log 2/\gamma)$ for 2, 3 and 4 competing cells.

bacterial division times of about 30 minutes, this result emphasizes the substantial impact that the choice of protein reporter could have on our ability to actively monitor and control cellular responses. To explore the effect that errors in the estimate of the absolute protein copy number would have on the success rate of the control law, we next assumed that measurements of the protein levels were corrupted by uncorrelated white noise with 20% standard deviation. Because our control law only needs to examine two cells (i.e., the chosen cell and the maximum of its competitors) and assign them to one of three coarse regions (see figure 3), the success rate is highly insensitive to errors in the measurement of the absolute protein numbers (see figure 6(a), dashed line).

In addition to the effects of time delays and measurement errors, we also consider the effects that extrinsic variability in protein synthesis rate would have on the control law effectiveness. In a series of simulations, we considered the case of variability in the synthesis rate, introduced by scaling both k_0 and k_1 by a common factor chosen from a normal distribution with a 10% standard deviation at the beginning of each independent simulation. To explore the effects that biases may have on the single-cell controllability, we assign the production rate of the chosen cell as the mean, the mean plus one standard deviation or the mean minus one standard deviation of the whole population. In figure 7(a) we show the average effectiveness of the control law based on the maximum protein expression in other cells. From the figure, we see that extrinsic variability can have a substantial effect on the average cell ranking. If the chosen cell has weaker production strength than its neighbors, then its uncontrolled ranking will be lower than the 50$^{\text{th}}$ percentile, but the addition of control can significantly increase this ranking (compare solid and dashed purple lines in figure 7(a)). If the chosen cell has stronger production rate than its competitors, then control can further increase that cell's advantage, at least for small populations size of about 70 cells or less (compare solid and dashed orange lines in figure 7(a)). We note that extrinsic variability plays a substantial role in the effectiveness of the control strategy from one random population to the next. To illustrate this concern, figure 7(b) plots the success probability for two cells with different relative production rates. When the chosen cell has an elevated production rate, control becomes very easy, but if the chosen cell's production rate falls below the threshold needed to maintain bistability (i.e., to the left of the red bar in figure 7(b)), successful control becomes almost impossible. We next examined 200 randomly generated populations of 30 cells apiece, where the chosen cell's transcription rate is the mean of transcription rates in other cells. Figure 7(c) plots the distribution of the success probability for the controlled (green) and uncontrolled (blue) system. From the figure, it is clear that the control law substantially improves the cell rankings overall, but for populations of heterogeneous cells, the specific success rate depend heavily on the particular population.

The control law we used thus far was always based on the maximum protein expression in other cells–the decision about the level of UV applied was made after assessing the maximum number of proteins in all other cells. We next explored how the effectiveness of control changes if the decision about the UV level is based on different statistics of the protein expression levels. Figure 8 shows the success rates versus the number of cells for additional control laws based upon the mean, the median, and the 75th percentile of the

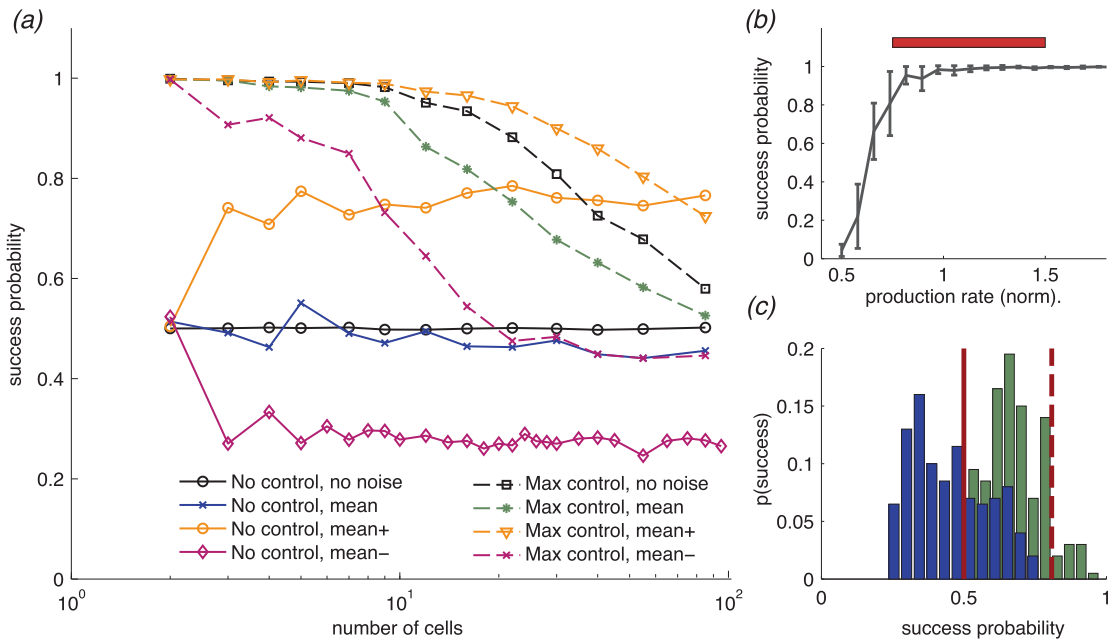

Figure 7. Influence of extrinsic variability in protein synthesis rates on the control law effectiveness. (a) Success probability as a function of the number of cells for different cases of variability of the synthesis rates. All cells are perturbed by 10% variability in their synthesis rate, and the synthesis rate of the chosen cell is fixed as the mean of the population (mean) plus or minus one standard deviation (mean+ and mean-, respectively). Results are shown for the uncontrolled system (no control), and for the system subjected to the control law based on the maximum protein expression in other cells (max control). For comparison, previous cases without extrinsic parameter variation are replotted (no noise). (b) Success probability for two cells versus the protein synthesis rate of the chosen cell divided by the synthesis rate of the other cell. The red bar on top indicates the regime for which a cell still exhibits bistability. Error bars indicate standard deviation computed over 20 simulations. (c) Distribution of the success probability for 200 simulations of different 30-cell populations with 10% random deviations in synthesis rates for the uncontrolled system (blue) and with the maximum control law (green). In all situations, the transcription rate of the chosen cell is equal to the population mean. We also plot reference lines for the system without protein synthesis noise (red lines, solid for the uncontrolled system, dashed for controlled system).

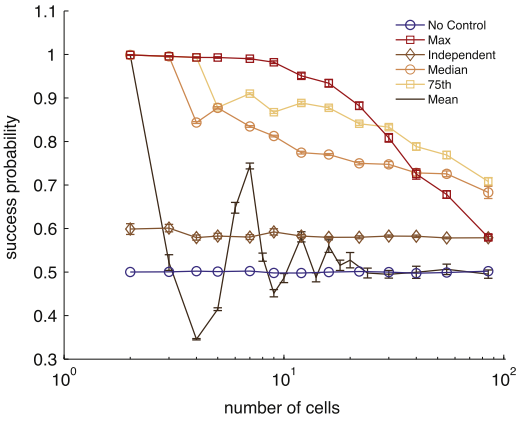

Figure 8. Comparison of the effectiveness of different control laws. Shown is the success probability as a function of the number of cells for different control strategies: the control law based on the maximum expression of other cells (red squares), the control law based on the mean expression, (black diamonds), the median (orange circle), the 75th percentile (gold squares) and the control law that only considers the chosen cell (brown diamonds). We plot for reference the uncontrolled system, for which the UV level is constant (blue circles). Error bars indicate the standard error of the mean (SEM) for 12 independent repeats.

neighboring cells. We also considered an independent control law where the UV depends only upon the level of protein in the chosen cell compared to a null cell assumed to have no expression (i.e., lying on the x-axis of figure 3). The success probabilities plotted in figure 8 show that our preliminary control law based on the maximum works the best for small populations. As the number of cells increases, the law based on the 75th percentile and then median begin to outperform the control based upon the maximum. The success probability for the independent control law does not vary with the number of cells, suggesting that for very large populations, such a controller may be optimal. Interestingly, the success probability for the mean is not monotonic. This is likely due to fact that the average of a finite population fluctuating within a bimodal probability landscape can be very different from the instantaneous state of any individual within that population.

The final extension of the analysis is to consider the genetic toggle switch model with two protein species. In this case our aim is to find a UV radiation control law, such that the expression of LacI is high in one cell and small in all others. As before, the control law will be UV radiation applied at three levels only and the choice depends on the current state of the system. However, in this case, the control law is based upon a limited amount of directly observable information and directly controllable dynamics. Specifically, we assume that only the level of LacI can be measured experimentally through use of a fluorescent protein

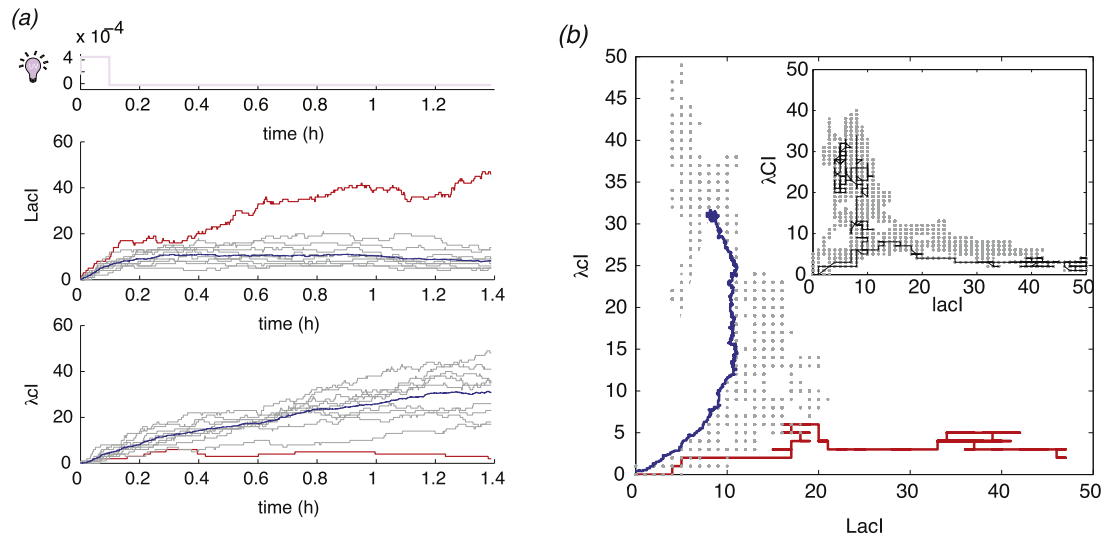

Figure 9. (a) Time evolution of the number of LacI proteins in ten cells expressing LacI and λcI proteins, subjected to UV radiation control law discussed in the main text. The trajectory for the chosen cell, in which we want to enhance LacI expression, is plotted in red. The trajectories for the nine other cells, in which we want to damp LacI expression, are plotted in grey and their mean is plotted in blue. (b) Number of LacI (x-axis) and λcI (y-axis) proteins in ten cells, as in panel (a). Parameters of the gene regulatory circuit (see main text) are fixed and identical for all cells. The inset in panel (b) shows two example trajectories of LacI and λcI content in two cells subjected to constant UV, under which the cell exhibits either high LacI or high λcI.

reporter, yet only the degradation rate for λcI can be controlled via application of UV radiation [20].

Given these constraints, we pose a control law that is essentially the reverse of the law we used in the previous analyses (compare to figure 3): high UV is applied when the system is in a state of low LacI expression in all cells, moderate UV is applied if the system has high LacI expression in the chosen cell and low expression in other cells (desired states), and low UV is applied if the system has high LacI expression in other cells, regardless of its expression in the chosen cell. Note that here we steer the system to the desired states indirectly, since the UV does not affect LacI expression directly and only the inhibitory interactions of LacI and λcI enable tuning of the LacI expression using UV radiation.

Figure 9 shows the results of the control law described here, applied to a population of ten cells. In the chosen cell (red curve) we drive the expression of LacI to be high (panel (*a*)), while the levels of λcI are driven to be low (panel (*b*)). For the other nine cells in the population, we drive the cells to the opposite behavior (grey curves represent individual trajectories and their mean is plotted in blue). To illustrate how dramatically different the controlled cell's behavior is in comparison to the others, panel (*c*) illustrates the same trajectories for the ten different cells, all on the LacI-λcI plane.

4. Discussion

Intrinsic stochasticity of biomolecular processes that involve small numbers of important molecules, creates uncertain fluctuations (or 'noise') that restrict our ability to make precise and accurate predictions for the behavior of individual cells. Conversely, for deterministic processes that lack these intrinsic uncertainties, knowledge of the initial conditions and reaction dynamics provide accurate and complete predictions for the system evolution in time. Since the ability to control behavior stems from the ability to predict that behavior under different input signals, the uncertainty underlying gene regulation suggests an obstacle that could make it more difficult to control individual cells without direct intervention at each individual cell. In this work, we found the opposite; stochasticity can break symmetry and actually enable precise control of many individual cells using a single input.

We illustrated that one could in principle exploit stochastic fluctuations to drive individual, yet identical cells each into a pre-specified, desired state. This capability was illustrated using two models of cell populations, where dynamics are governed by simplified versions of the genetic toggle switch (figure 1(a), (b)) [18, 19]. We asked if application of a single universal input (e.g., UV radiation) applied simultaneously to the whole population of cells could force one chosen cell to exhibit a qualitatively different phenotype than the others. For a deterministically varying process, we showed that the phenotype of each cell depends entirely on its initial condition, and therefore it would be impossible to individually control cell phenotypes without first specifying these initial conditions. Since it is well recognized that many single-cell gene regulation behaviors can be understood better using stochastic formulations [7], we also explored the case where the protein levels could fluctuate according to discrete stochastic reactions. We found that these fluctuations render identical cells to become distinct

and susceptible to control inputs. In turn, this fluctuating susceptibility allowed for formulation of control laws that depend on the observed and continually perturbed states of the system.

Using a combination of stochastic simulations and direct solutions of the chemical master equation, we designed a simple control law (see figure 3) that drives many cells, with high probability, to pre-specified states of high protein expression in one specific cell and low protein expression in the rest. We applied this control law to two different versions of the genetic toggle switch for different populations sizes, and we also explored how time delays inherent to the slow-folding dynamics of fluorescent proteins would affect our ability to control cellular phenotypes. Our analyses showed the principle that a simple strategy could successfully control small populations provided that the time delay is smaller than the half life of the protein whose degradation is under control (see figure 6). We also showed that the chosen control law could work in the presence of 20% or more measurement uncertainties (see figure 6(a)) as well as for 10% extrinsic variabilities in single-cell production rate parameters (see figure 8).

The control law we used in this preliminary analysis permitted exactly three discrete levels of UV radiation. For two cells, the UV level was chosen based upon a simple comparison, and for N cells, the UV level was chosen based on the protein level in the chosen cell and cells at different percentile ranked within the remaining population. In all cases within this study, a single non-optimized control strategy has been used (see figure 3). Controlling the UV level as a continuous function of the two cell's protein levels would likely lead to far more effective results, but could also lead to increased sensitivity to measurement errors. Furthermore, for populations of heterogeneous cells, our preliminary comparison of the chosen cell with a summary statistic of the other cells is an arbitrary choice meant to illustrate the feasibility of control. In principle one could specify the UV level in terms of the any other function of the N dimensional levels if single-cell expression. It is extremely likely that improved control strategies could drive the system to the desired state with far greater success probability, and the optimization of such multi-cell stochastic control strategies represents a rich area for future exploration. Our computationally demonstrated feasibility for the use of noise to control individual phenotypes of multiple cells with a single input, suggests that noise-enhanced control could eventually play an important role in biomedical engineering. Such approaches could complement genetic techniques of synthetic biology and the single-cell perturbation techniques of microfluidics and optical control to improve capabilities for multi-cellular control in future tissue engineering and regenerative medicine applications.

Acknowledgments

This study began as a project at the 2014 q-bio Summer School (qbSS, Albuquerque, NM) and was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R25GM105608. BM thanks all of the lecturers and students of the 2014 qbSS for their valuable feedback and discussions related to this project. PS was supported by the EU through the European Social Fund, contract number UDA-POKL.04.01.01–00-072/09–00. NG has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement no. [338200].

References

[1] Milias-Argeitis A, Summers S, Stewart-Ornstein J, Zuleta I, Pincus D, El-Samad H, Khammash M and Lygeros J 2011 *Nat. Biotechnol.* **29** 1114–6
[2] Uhlendorf J, Miermont A, Delaveau T, Charvin G, Fages F, Bottani S, Batt G and Hersen P 2012 *Proc. Natl. Acad. Sci. USA* **109** 14271–6
[3] Blake W J, Kræn M, Cantor C R and Collins J J 2003 *Nature* **422** 633–7
[4] Raser J M and O'Shea E K 2005 *Science* **309** 2010–3
[5] Eldar A and Elowitz M B 2010 *Nature* **467** 167–73
[6] Raj A and van Oudenaarden A 2008 *Cell* **135** 216–26
[7] Munsky B, Neuert G and van Oudenaarden A 2012 *Science* **336** 183–7
[8] Kræn M, Elston T C, Blake W J and Collins J J 2005 *Nat. Rev. Gen.* **6** 451–64
[9] Singh A and Soltani M 2013 *PLOS ONE* **8** e84301
[10] Losick R and Desplan C 2008 *Science* **320** 65–68
[11] Veening J W, Smits W K and Kuipers O P 2008 *Annu. Rev. Microbiol.* **62** 193–210
[12] Arkin A, Ross J and McAdams H H 1998 *Genetics* **149** 1633–48 (PMID: 9691025)
[13] Weinberger L S, Burnett J C, Toettcher J E, Arkin A P and Schaffer D V 2005 *Cell* **122** 169–82
[14] Weinberger L S, Dar R D and Simpson M L 2008 *Nat. Genet.* **40** 466–70
[15] Hasty J, Pradines J, Dolnik M and Collins J J 2000 *Proc. Natl. Acad. Sci. USA* **97** 2075–80
[16] Kussell E and Leibler S 2005 *Science* **309** 2075–8
[17] Balaban N Q, Merrin J, Chait R, Kowalik L and Leibler S 2004 *Science* **305** 1622–5
[18] Gardner T S, Cantor C R and Collins J J 2000 *Nature* **403** 339–42
[19] Kobayashi H, Kræn M, Araki M, Chung K, Gardner T S, Cantor C R and Collins J J 2004 *Proc. Natl. Acad. Sci. USA* **101** 8414–9
[20] Munsky B and Khammash M 2010 *IET. Sys. Biol.* **4** 356–66
[21] Warren P B and ten Wolde P R 2005 *J Phys. Chem.* B **109** 6812–23
[22] Lipshtat A, Loinger A, Balaban N Q and Biham O 2006 *Phys. Rev. Lett.* **96** 188101
[23] Shimizu-Sato S, Huq E, Tepperman J M and Quail P H 2002 *Nat. Biotechnol.* **20** 1041–4
[24] Levskaya A, Weiner O D, Lim W A and Voigt C A 2009 *Nature* **461** 997–1001
[25] Kaufmann S H and Earnshaw W C 2000 *Exp. Cell. Res.* **256** 42–49
[26] Ptashne M 2004 *A Genetic Switch: Phage Lambda Revisited* vol 3 (New York: Cold Spring Harbor Laboratory Press)
[27] Gillespie D T 1977 *J. Phys. Chem.* **81** 2340–61

[28] Munsky B and Khammash M 2006 *J. Chem. Phys.* **124** 044104
[29] Sidje R B 1998 *ACM Trans. Math. Softw.* **24** 130–56 http://doi.acm.org/10.1145/285861.285868
[30] Burrage K, Hegland M, Macnamara S and Sidje R 2006 *Proc. of The A. A. Markov 150th Anniversary Meeting* pp 21–37
[31] MATLAB 2014 *version 8.3.0.532 (R2014a)* Natick, Massachusetts, The MathWorks Inc.
[32] Munsky B 2012 Modeling cellular variability *Quantitative Biology From Molecular to Cellular Systems* ed M E Wall (New York: CRC Press) pp 233–66
[33] Albeck J G, Burke J M, Spencer S L, Lauffenburger D A and Sorger P K 2008 *PLoS Biol.* **6** e299
[34] Spencer S L and Sorger P K 2011 *Cell* **144** 926–39
[35] Shaner N C, Steinbach P A and Tsien R Y 2005 *Nat. Methods.* **2** 905–9
[36] Nagai T, Ibata K, Park E S, Kubota M, Mikoshiba K and Miyawaki A 2002 *Nat. Biotechnol.* **20** 87–90
[37] Campbell R E, Tour O, Palmer A E, Steinbach P A, Baird G S, Zacharias D A and Tsien R Y 2002 *Proc. Natl. Acad. Sci. USA* **99** 7877–82