Exploiting noise, non-linearity, and feedback for differential control of multiple synthetic cells with a single optogenetic input

Michael P May[†] and Brian Munsky[*,†,‡]

†*School of Biomedical Engineering, Colorado State University, Fort Collins, CO, USA, 80523*

‡*Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO, USA, 80523*

E-mail: brian.munsky@colostate.edu

Abstract

Synthetic biology seeks to develop modular bio-circuits that combine to produce complex, controllable behaviors. These designs are often subject to noisy fluctuations and uncertainties, and most modern synthetic biology design processes have focused to create robust components to mitigate the noise of gene expression and reduce the heterogeneity of single-cell responses. However, deeper understanding of noise can achieve control goals that would otherwise be impossible. We explore how an "Optogenetic Maxwell Demon" could selectively amplify noise to control multiple cells using single-input-multiple-output (SIMO) feedback. Using data-constrained stochastic model simulations and theory, we show how an appropriately selected stochastic SIMO controller can drive multiple different cells to different user-specified configurations irrespective of initial condition. We explore how controllability depends on cells' regulatory structures, the amount of information available to the controller, and

the accuracy of the model used. Our results suggest that gene regulation noise, when combined with optogenetic feedback and non-linear biochemical auto-regulation, can achieve synergy to enable precise control of complex stochastic processes.

Keywords: *synthetic biology, autoregulation, gene regulation noise, optogenetic feedback control, single-input-multiple-output control, Maxwell's Demon*

Introduction

Synthetic biology seeks to develop and characterize biological circuits and modular components that can be reliably re-engineered, re-assembled, and controlled to produce complex biological behaviors.[1] The design and implementation of modular components as simple functional units[2] capable of being assembled to perform desired regulatory needs has yielded powerful capabilities of synthetic cells to perform specific actions in response to specific stimuli.[3] Early advances led to the development of programmed cells that are capable of complex logic like switching, self-regulation, and fast acting control.[4] Genetically engineered switches were originally built in bacteria,[5] but have been extended to yeast,[6] mammalian cells,[7] and even multicellular plants.[8] In turn, these simple switches have led to more complex engineered biological systems and biotechnologies capable of performing tasks like controlling cells to behave as digital displays.[9]

Much of the design process for synthetic biology has focused directly on building better biological components, such as creating more sophisticated gene regulatory structures,[10–12] introducing new response elements or reporters,[13] or introducing more orthogonal cellular signals.[14] Development work on gene regulatory structure has led to substantial advances in phenotype control in plant biology and targeted or modified protein turnover in therapeutics.[15,16] Improvements to response reporters and the experimental techniques used to analyze cells, especially in the form of fluorescent protein reporters ,[17,18] real time single-gene MCP-MS2-based transcription elongation assays,[19] and nascent chain translation elongation assays[13,20–22] has made it possible

for cells to transmit their internal states to human observers. These technologies allow for more observation and control, not just at the protein level, but at the gene and RNA levels as well. Advances in cellular signals have also introduced the potential for synthetic regulatory modules in separate cells to communicate with one another and control multicellular population dynamics.[23–25] For example, by considering a simple model for the effect of cellular quorum sensing on cell densities, simple circuits can be tuned to control cell densities.[26]

Although the above advances have been developed primarily to control autonomous biological behaviors, these improvements to regulatory structures, response reporters, and signaling capabilities can also provide a framework to allow observers or external electronic devices to monitor cellular environments and dynamically reprogram the cellular logic. When coupled to advances in microfluidics, these capabilities introduce a new part-biology-part-machine (or cyber-organic[27]) paradigm that adds new possibilities for distributed external and internal biological control of synthetic biological circuits.[28] In particular, recent developments in optogenetics[29,30] have greatly increased the speed and sensitivity by which external signals can be communicated from humans or machines to cells. Using these advances in microfluidics and light-activated gene regulatory elements is rapidly improving the potential to integrate carbon- and silicon-based circuits, which in turn is making hybrid bio-electronic circuits far more powerful than before.

A key challenge to integrating cell-based genetic circuity with electronic control is that an uncountable number chemical species and regulating bodies must diffuse and interact with one another in space and time within each cell. Tractable analysis requires immense simplifications of these infinitely complex and chaotic dynamics, and such simplifications naturally result in large uncertainties that can only be accounted for through the introduction of approximate models with stochastic analyses. One of the greatest challenges to improving externally controlled cellular behaviors is that this 'noise' in gene regulation introduces large amounts of single-cell heterogeneity which must accounted for.[31–33] Under the context of this noise, cellular responses

are probabilistic–their distributions may shift gradually and even be statistically predictable under environmental or genetic manipulations,[34–36] but individual cells appear to behave at random with very little information about their instantaneous external environment.[37] Unfortunately, this noise makes it difficult to precisely predict how or when an individual cell will respond to a new environmental stimulus. Although recent work has shown that external feedback can control and reduce cellular heterogeneity within a large population,[38] it may seem unlikely that any feedback control strategy could reliably guarantee that specific cells within a population will respond as desired, and independently of their initial conditions.

Most efforts on external feedback control in synthetic biology have focused on the use of chemical or optical inputs to manipulate cell population averages[39] or to control individual cells within a larger population.[39] Such efforts can be classified as single-input-single-output (SISO) or multiple-input-single-output (MISO) control in that they seek only to push cells to a single phenotype. For example, recent experimental and computational studies[40–42] have used periodic chemical input fluctuations to control one cell or a population of multiple cells to be as close as possible to the same unstable fixed point, a task that is similar to the inverted-pendulum problem in classical control theory. Other recent work has sought to control multiple individual cells, each with their own tailored optogenetic inputs,[29] which corresponds to multi-input-multi-output control (MIMO). SISO and MISO control are limited to control only a single cellular response at a time, while MIMO requires advanced hardware such as digital micro-mirror devices that devote a separate input to each individual cell.[29,43] However, the combination of synthetic biological designs, precise external controls, and quantitative measurements and models of single-cell noise could create new opportunities for single-input-multiple-output (SIMO) control, where multiple individual cells could be controlled to achieve *different phenotypes*, but requiring only a single input signal. In this article, we explore the potential of a noise-enabled SIMO control strategy that is similar to a hypothetical Maxwell's Demon (MD), who watches the random process and identifies short instances in time when specific cells have randomly increased sen-

sitivities to small perturbations.[44] In the context of synthetic biology, we explore how advances in fast fluorescent reporters allow this MD to watch biological responses; how noise breaks the symmetry between identical cells in identical environments; and how advances in optogenetics may enable the MD to drive specific cells toward specific phenotypes, even when each cell always experiences the exact same input signal as every other cell.

Through simulation of multiple models that have been parametrized from existing bulk level optogenetic control experiments, we demonstrate that realizing an optogenetic MD for use in genetic regulation applications requires not only careful consideration of the cellular regulatory systems to be controlled, the fluorescent sensors to be observed, and the optogenetic inputs to be delivered, but it is also necessary to build predictive stochastic models and deterministic SIMO control algorithms to serve as its brain. In this article, we use a combination of simulations and theoretical analyses to explore how existing biological parts could be combined with models in the context of optogenetics to control the gene expression of multiple cells at once, even when both cells receive the same light signal at the same time and have the exact same genetics. We then introduce a new probabilistic model predictive control (pMPC) strategy that can in principle control multiple cells to different phenotypes, even when only observing a single cell. Finally, we show that controls that are designed and optimized using one approximate model of the biological system can be effective to control a much more complex biological process whose mechanisms are not considered during the controller design.

Results and Discussion

We start by defining a set of two models, each with a different level of complexity, to describe the dynamics of optogenetically activated T7 polymerase in temporally-varying light conditions. We then use experimental data from Baumschlager et al[45] to independently constrain each of these models to the same data. We then take the

simpler of the two models and extend it to include a typical auto-regulation module, and we examine the performance of this extended model under different fluctuating input signals at different frequencies using deterministic and stochastic analyses. We then show how intrinsic noise in the system dynamics can be utilized by a feedback controller to break symmetry in the process and force a system of two cells each to obtain specified phenotypes and independent of initial conditions. We then propose a new probabilistic model predictive controller scheme that is capable to differentially control multiple cells even when only one is directly observed. Finally, we demonstrate in principle that an optogenetic controller that is identified using a coarse-grained simplified model is capable to control behavior of a more complicated system with different and hidden dynamics.

A data-constrained model for light-induced T7 polymerase activation and gene expression

We begin by developing an *unregulated* model (denoted as $\mathcal{M}_U$) to describe the light-induced activation of an optogenetically controlled T7 polymerase as studied in Baumschlager et al.[45] As depicted in Fig. 1A, this system contains two light activated T7 domains ($T7_n$ and $T7_c$) that are produced at constant rates k_n and k_c, and which degrade in a first order decay process with rate γ_M. These domains dimerize when subject to light activation leading them to form an active T7 polymerase at a light dependent rate of $u(\phi)$ and the complex dissociates at a rate of k_{i1} and decays at a rate γ_T. The T7 polymerase dimer can bind to, or unbind from, the gene at rates of k_{f2} and k_{i2}, respectively. When bound, active protein production occurs at a rate of k_{f1}, where transcription and translation are lumped into a single event. Proteins are assumed to degrade according a first order rate process with rate γ_P. These interactions are described by the following reactions

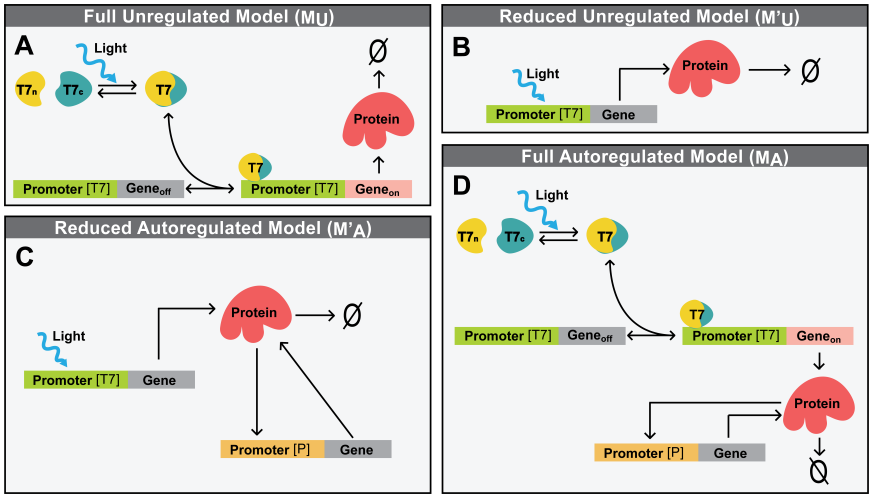

Figure 1: Diagrams of models for light-induced gene regulation with and without auto-regulation. (A) The full physical model ($\mathcal{M}_U$) describes the full mechanistic processes of T7 polymerase dimerization and gene activation in addition to protein production and degradation. (B) A simplified model ($\mathcal{M}'_U$) lumps together all T7 dimerization and gene activation dynamics so that protein is produced at a light-controlled rate. (C) The simplified model is extended ($\mathcal{M}'_A$) to include auto-regulation through the addition of a secondary promoter. (D) The full mechanistic model is also extended ($\mathcal{M}_A$) through inclusion of the same auto-regulation promoter.

$$\mathcal{M}_U = \left\{ \begin{array}{rcl} \emptyset & \underset{k_n}{\overset{\gamma_M[T7_n]}{\rightleftharpoons}} & T7_n, \\ \emptyset & \underset{k_c}{\overset{\gamma_M[T7_c]}{\rightleftharpoons}} & T7_c, \\ T7_n + T7_c & \underset{u(\phi)[T7_n][T7_c]}{\overset{k_{i1}[T7]}{\rightleftharpoons}} & T7\ , \\ T7 & \xrightarrow{\gamma_T[T7]} & \emptyset\ , \\ T7 + g_{\text{off}} & \underset{k_{f2}[T7][g_{\text{off}}]}{\overset{k_{i2}[g_{\text{on}}]}{\rightleftharpoons}} & g_{\text{on}}\ , \\ g_{\text{on}} & \xrightarrow{k_{f1}[g_{\text{on}}]} & P + g_{\text{on}}\ , \\ P & \xrightarrow{\gamma_P[P]} & \emptyset\ , \end{array} \right\} \tag{1}$$

where the first two bidirectional reactions describe production and decay of $T7_n$ and $T7_c$; the third bidirectional reaction describes light-induced reversible dimerization, where the light induction level is denoted as ϕ and its effect is modeled by the function $u(\phi)$; the fourth unidirectional reaction describes the decay of the T7 dimer; the fifth

bidirectional reaction describes the T7 association and dissociation to the gene; and final two unidirectional reactions describes the production and degradation for the resulting protein product, where transcription and translation have been lumped into a single reaction. The rate for each reaction is given directly above or below its respective arrow.

The ODE for $\mathcal{M}_U$ can be written in vector form as

$$\frac{d}{dt}\begin{bmatrix} [T7_n] \\ [T7_c] \\ [T7] \\ [g_{\text{on}}] \\ [P] \end{bmatrix} = \begin{bmatrix} -u(\phi)[T7_n][T7_c] + k_{i1}[T7] + k_n - \gamma_M[T7_n] \\ -[T_7n][T7_c] + k_{i1}[T7] + k_c - \gamma_T[T7_c] \\ u(\phi)[T7_n][T7_c] - k_{i1}[T7] - k_{f2}[T7](g_{\text{total}} - [g_{\text{on}}]) + k_{i2}[g_{\text{on}}] - \gamma_T[T7] \\ k_{f2}[T7](g_{\text{total}} - [g_{\text{on}}]) - k_{i2}[g_{\text{on}}] \\ k_{f1}[g_{\text{on}}] - \gamma_P[P] \end{bmatrix}, \tag{2}$$

where we have assumed a fixed number of gene copies $[g_{\text{total}}] = [g_{\text{on}}] + [g_{\text{off}}]$ to remove the variable $[g_{\text{off}}]$ from the equations. Once written in this form, Model $\mathcal{M}_U$ can be integrated numerically for any given set of initial conditions and parameters. The model parameters were then fit to the experimental data from Baumschlager et al,[45] in which the system was subjected to three different levels of UV radiation:

$$\phi(t) = \begin{cases} 320 \text{ Watts/cm}^2, \text{ for } 0 \leq t < 270 \text{ minutes}, \\ 0 \text{ Watts/cm}^2, \text{ for } 270 \leq t < 570 \text{ minutes}, \\ 20 \text{ Watts/cm}^2, \text{ for } 570 \leq t \text{ minutes}. \end{cases}$$

The parameters of $\mathcal{M}_U$ and $\{u(\phi_1), u(\phi_2), u(\phi_3)\}$ are simultaneously fit to the measured time series fluorescent protein trajectory. This fit suggests u=[0.4060, 0.00, 0.0400] molecules^{-1}min^{-1} when ϕ=[320, 0, 20] Watts/cm^2, respectively. To interpolate for intermediate values of light intensity, the function $u(\phi)$ is then defined as the cubic spline of the three data points and is shown with a red line in Fig. 2B. The resulting fit of the model to the data is shown by the red dashed lines in Fig. 2A, and the remaining parameters of the model fit are shown in Table 1.

Table 1: Parameters of the full model system.

Parameter Name	Parameter Value	Units
k_n	2.00×10^{-1}	molecules/min
k_c	6.00×10^{-1}	molecules/min
k_{i1}	2.00×10^{2}	min^{-1}
γ_M	5.00×10^{-2}	min^{-1}
γ_T	5.00×10^{-2}	min^{-1}
k_{i2}	5.00×10^{-1}	min^{-1}
k_{f2}	5.00×10^{-1}	molecules/min
k_{f1}	1.42×10^{-2}	min^{-1}
γ_P	2.03×10^{-2}	min^{-1}

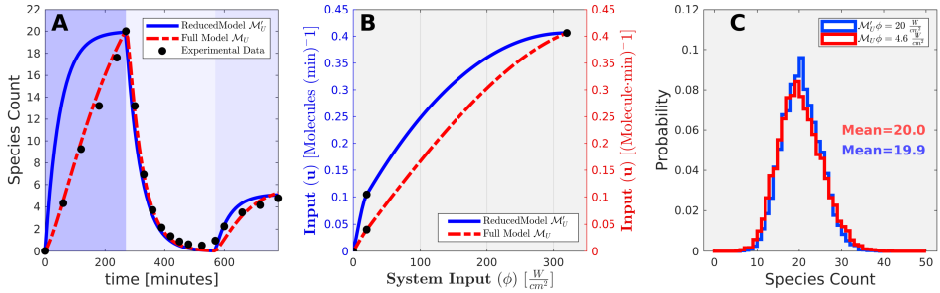

Figure 2: Parameter estimation of full and reduced unregulated models, $\mathcal{M}_U$ and $\mathcal{M}'_U$, respectively. (A) Fits of $\mathcal{M}_U$ (red line) and $\mathcal{M}'_U$ (blue line) to experimental data from Baumschlager et al.[45] (black dots).(B) Calibration curves representing the conversion of light [Watts per cm^2] to the associated reaction rates for $\mathcal{M}_U$ (red) and $\mathcal{M}'_U$ (blue). (C) Steady state histograms predicted by models $\mathcal{M}_U$ (red) and $\mathcal{M}'_U$ at inputs that are calibrated to results in an average expression of 20 molecules per cell.

Having determined a baseline ODE-based model that yields a good fit to existing experimental data for the system's temporal response, in the next sections we will specify a much simpler, but less accurate, version of this model and use that approximate model to perform stochastic analyses, suggest design modifications, and specify a controller that can drive differential gene expression among two or more cells using a single external input signal.

A substantially reduced and approximate model for T7 activation and gene expression

Although we were able to find many good parameter sets so that model $\mathcal{M}_U$ could match the data,[45] identification of a single unique 'best' parameter set is infeasible due to the high number of parameters, severe limitations on available experimental data, and sloppiness[46] in the model parameters. Finding a simpler, but better constrained model would not only help to reduce sensitivity to unknown parameters, but could also dramatically reduce computational costs when using that model for design decisions or for the specification of feedback control strategies. To simplify these model reactions and to obtain a more identifiable parameter set, we next propose a simple generalized birth-death model in which the protein production rate is given by $k_0+u'(\phi)$, where k_0 is the baseline production rate with no light input, and $u'(\phi)$ is a light-dependent control input function. We use the apostrophe notation $(\cdot)'$ to denote use of the reduced model in which the units of the control signal at a given light intensity have been adjusted to molecules per minute. Under this simple rule and assuming linear decay at rate γ_P, the approximate dynamics of $P(t)$ are written simply as:

$$\frac{dP}{dt} = k_0 + u'(\phi) - \gamma_P P. \tag{3}$$

In this model, which we denote as $\mathcal{M}'_U$, the parameters k_0 and γ_P and the specific values of $u'(\phi)$ at $\phi \in \{320, 0, 20\}$ Watts/cm^2 are again simultaneously fit to capture the time dynamics of the experimental data.[45] This fit suggests $u'(\phi)$=[0.4060, 0.00, 0.1044] molecules/min when ϕ=[320, 0, 20] Watts/cm^2 respectively, and the cubic spline of these three data points yields the calibration curve $u'(\phi)$ as shown in Fig. 2B (blue line). The resulting fit of model $\mathcal{M}'_U$ is shown in the solid blue line Fig. 2A, and the remaining parameters for the reduced model are shown in Table 2.

We next extended both the original model and the simplified model to include dis-

Table 2: Parameters of the reduced model system.

Parameter Name	Parameter Value	units
γ_P	2.03×10^{-2}	1/min
k_0	1.00×10^{-4}	molecules/min

crete stochastic events for protein production and degradation, as well as for T7 dynamics for the full model. Using the exact same parameter values as for the ODE analysis, we then simulated the model using the Stochastic Simulation Algorithm (SSA,[47]). Fig. 2C shows the probability distribution of P sampled using 10^5 independent SSA runs, each simulated to 3000 minutes and only using the last data point of each simulation for either the full model (red) or the simplified model (blue). Both models provide reasonably good, but not identical, matches to the mean protein levels and the overall time scales as shown in Fig. 2A, and we observed strong agreement in their stationary distributions. However, it remains to be seen if the differences in architecture and time scales need to be accounted for when we use the simplified model ($\mathcal{M}'_U$) to guide our modification of the gene regulatory system and to design a feedback controller that remains effective even when applied to the more complex original model ($\mathcal{M}_U$).

Addition of an autoregulation motif creates light-dependent bistability

Starting from the simplified model, $\mathcal{M}'_U$, we next asked if a common auto-regulation motif could be added to introduce light-dependent bistability for the system and so that we could explore how that added motif would impact the controllability of the overall system. Fig. 1C shows a schematic of the new model denoted as $\mathcal{M}'_A$, which has auto-regulation due to the addition of a secondary promoter that is self-activated by protein P to produce more of itself. To incorporate this auto-regulation behavior, a Hill function production rate is assumed with high cooperativity, and the new rate equation becomes:

$$\frac{dP}{dt} = u'(\phi) + k_0 + \kappa \frac{P^\alpha}{P^\alpha + \beta^\alpha} - \gamma P, \tag{4}$$

where the first term $u'(\phi)$ corresponds to the control input as a function of light input; the second and third terms correspond to the Hill function activity of the auto-regulation promoter with leakiness k_0, and the final term corresponds to the first order decay of the protein. The Hill parameters (β, α) of $\mathcal{M}'_A$ are chosen in order to exhibit bi-modal behavior in the dynamics of P. All parameters of this auto-regulatory model are shown in Table 3.

Table 3: Hill function parameters for auto-regulation motif.

Parameter Name	Parameter Value	Units
κ	4.06×10^{-1}	molecules/min
γ	2.03×10^{-2}	min^{-1}
β	20	molecules
α	8	unitless
k_0	1×10^{-4}	molecules/min

Fig. 3A shows bifurcation diagrams for the simplified unregulated and auto-regulated models, $\mathcal{M}'_U$ and $\mathcal{M}'_A$, respectively. These diagrams show that as the light input sweeps *slowly* from zero to 0.5 Watts per cm^2, bifurcation and hysteresis become apparent in the auto-regulated model ($\mathcal{M}'_A$, light/dark blue), but these effects are not observed in the unregulated model ($\mathcal{M}'_U$, orange/red). For either model, low and high light inputs each result in a single stable point at low or high expression, respectively. For intermediate light inputs, however, two history-dependent stable points coexist for $\mathcal{M}'_A$, and it is possible for two cells to maintain different stable points (or phenotypes) provided that the light intensity remains in the bi-stable region and that the cells begin in the separate basins of attraction for the different stable points.

In the hysteresis plots of Fig. 3A, it is assumed that the light input sweeps very slowly so that the response reaches equilibrium at each light level before subsequent changes. However, in the context of feedback control, we are more interested in how cells respond to light fluctuations at faster, transient time scales. Therefore, we next test the stability of input-to-output behaviors under time-varying inputs. We start

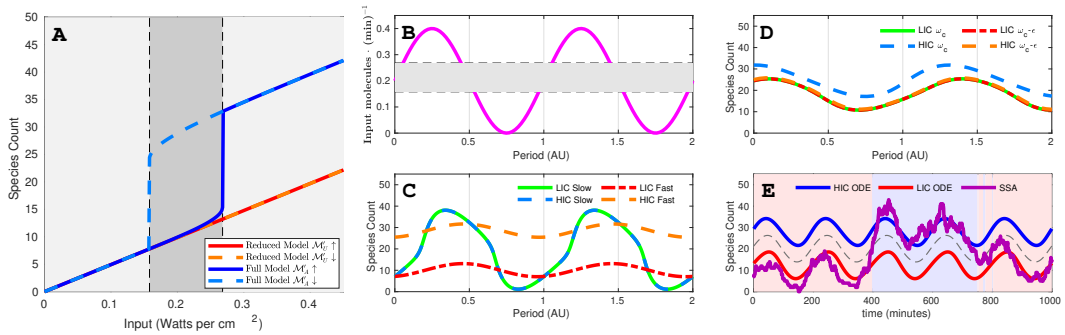

Figure 3: Input-output analysis for the simplified model with and without auto-regulation (Models $\mathcal{M}'_U$ and $\mathcal{M}'_A$). (A) Bifurcation (hysteresis) analyses of $\mathcal{M}'_U$ (red shades) and $\mathcal{M}'_A$ (blue shades). Solid (dashed) lines depict the change in the steady state solution as UV intensity is increased (decreased). Bi-modality (i.e., two possible steady states at the same control input) and hysteresis (i.e., different paths of solid and dashed lines) only occur for $\mathcal{M}'_A$. (B) Sinusoidal input used for temporal excitation shown for two periods. (C) Steady state temporal behavior of $\mathcal{M}'_A$ under sinusoidal input shown for two periods. Dashed lines correspond to high initial conditions, and solid lines correspond to low initial conditions. Different colors correspond to different input signal frequencies: fast (0.01 RPM, red/orange), slow (0.001 RPM, blue shades). All plots show steady state temporal response after a transient time of at least two oscillation periods or 3500 minutes, whichever is longer. (D) Same as (C) but with frequencies 0.0043522 RPM (red/orange) and 0.0043525 RPM (blue shades). The $\mathcal{M}'_A$ maintains memory of initial condition provided that input frequency is greater than a critical value (i.e., solid and dashed lines remain distinct). (E) Capability of Model $\mathcal{M}'_A$ to track inputs assuming stochastic fluctuations as analyzed using an extended SSA[47] with extra reactions.[48]

simulations for two cells with identical parameters, but at different initial conditions (i.e., one at a high initial concentration of 40 molecules per cell and another at a low concentration of 0 molecules per cell), and we subject these both to the same sinusoidally-varying input signal whose range encompasses both bifurcation points as shown in Fig. 3B. Fig. 3C shows the steady state trajectories for two pairs of such systems, where dashed lines correspond to trajectories that start at low initial conditions and solid lines correspond to trajectories starting at high initial conditions. In all cases, the system is subject to at least four cycles or 3500 minutes so that transient dynamics have had time to decay, and the time axis is scaled to show the response over two oscillation periods. Fig. 3C (blue shades) shows that when the input frequency is slow ($\omega_{slow} = 0.001$ rotations per minute (RPM)), the system loses memory of its initial condition and the trajectories from both initial conditions decay to a single trajectory. However, when the frequency is fast (red and orange trajectories, $\omega_{fast} = 0.01RPM$), the system can maintain memory indefinitely. Fig. 3D shows that the cut off frequency for this maintenance of memory is sharp in that memory is possible at a frequency of ω_c (blue shades) but is lost at a slightly slower frequency of $\omega_c - \varepsilon$, where $\omega_c = 0.0043525$ (red/orange) RPM is the critical frequency and $\varepsilon = 3 \times 10^{-7}$ RPM is a small perturbation to that frequency.

Intrinsic noise can drive cells to switch phenotypes

Using deterministic analyses of the bistable model, we have seen that two cells with different initial conditions maintain separate phenotypes as they respond to the same fluctuating input signal. The flip side of this deterministic result is that two fully converged solutions of the same ODE never diverge, such that two cells starting at the exact same initial condition will never express unique phenotypes even when bi-stability is possible. However, low copy numbers of important regulatory molecules (DNA, RNA and proteins) often result in stochastic fluctuations in cellular concentrations (also known as 'intrinsic noise') that dramatically affects both of these results. When added to a bistable deterministic process, noise can drive two cells starting at the same initial

phenotype to diverge or even drive two cells to exchange phenotypes by chance over time.[49,50] With this possibility in mind, we next ask how noise would affect the ability of cells to track a temporally varying input signal. For this, we examined the production and degradation reactions and converted model $\mathcal{M}_A$ to an equivalent discrete stochastic model with the exact same rate parameters, and we explored how discrete intrinsic noise of stochastic models could be used drive cells to separate phenotypes. To extend the SSA to approximate time-varying inputs, we adopted an approach similar to that in Voliotis et al,[48] and added a fast 'null event' reaction that updates the clock and input signal value on a time scale that is much faster (i.e., average of 100 events per period of the input signal) than that of the input signal fluctuations. We then compared the ODE and the SSA analysis of $\mathcal{M}'_U$ under a sinusoidal input with moderate input frequency of $\omega = 5.00 \times 10^{-3}$ RPM $> \omega_c$ for which the ODE trajectories maintain memory of their initial conditions. Although the ODE solutions (smooth lines in Fig. 3E) will never converge, the stochastic trajectories (purple fluctuating trajectories) switch occasionally between the two fluctuating phenotypes. In other words, with the addition of noise to the system, each cell can slowly 'forget' its original configuration. Moreover, the probability of switching depends on the transient stochastic state of the process and the frequency and amplitude of external input fluctuations. Previous studies have observed similar effects for how noise creates variation in a population of cells, and past feedback control efforts have sought to counteract this variation to keep all cells at a chosen (and in some cases unstable) phenotype.[40–42] In the next section, we do not try to reduce variability among cells, but rather we seek to exploit the condition- and time-dependent disruption of symmetry to push one cell to a chosen phenotype while forcing another cell or group of cells toward a different chosen cell fate.

Finite State Projection analyses uncover effective strategies to differentially control two cells using a single input

We consider a system of N_c cells, each with identical regulatory mechanisms and parameters, but whose fluctuating protein concentrations at time t are denoted by $\tilde{x}_1(t), \tilde{x}_2(t), \ldots$, which we can arrange into the vector $\tilde{\mathbf{X}}(t) = [\tilde{x}_1(t), \tilde{x}_2(t), \ldots]^T$. Here, the notation $(\tilde{\cdot})$ denotes that the corresponding quantity (e.g., protein copy number) is the result of a stochastically fluctuating process. Our overall goal is to design a feedback control law to force $\tilde{\mathbf{X}}(t)$ as close as possible toward an arbitrary target state $\hat{\mathbf{T}}$. For an example with two cells, $\hat{\mathbf{T}} = [\hat{T}_1, \hat{T}_2]^T = [30, 10]^T$ would correspond to having the first cell in the high expression phenotype and the second cell in the low expression phenotype. In general, this fluctuating control signal could depend on measurements of $\tilde{\mathbf{X}}(t)$ and/or the current time according to some as yet to be determined control function $\tilde{u} = u(\tilde{\mathbf{X}}(t), t)$. We define a cost function as the expected squared Euclidean distance between $\tilde{\mathbf{X}}(t)$ and $\hat{\mathbf{T}}$ at steady-state, which can be written as

$$\begin{aligned} J &= \lim_{t\to\infty} \mathbb{E}\{|\tilde{\mathbf{X}}(t) - \hat{\mathbf{T}}|_2^2\} \\ &= \lim_{t\to\infty} \sum_{i,j,\ldots} P(\tilde{x}_1(t) = i, \tilde{x}_2(t) = j, \ldots) \Big((i - \hat{T}_1)^2 + (j - \hat{T}_2)^2) + \ldots \Big) \\ &= \lim_{t\to\infty} \sum_{i,j} P_{ij}(t) C_{ij} = \mathbf{C}\mathbf{P}_\infty, \end{aligned} \tag{5}$$

where $\mathbf{C}$ is a constant vector of squared Euclidean distances of each state from the target, and $\mathbf{P}_\infty$ is the steady-state probability mass vector (i.e., the stationary probability for each unique value of $\tilde{\mathbf{X}}$).

As described in Methods, the master equation for a finite number of cells subject to a fluctuating state- or time-dependent input signal can be written as

$$\frac{d}{dt}\mathbf{P} = (\mathbf{A}_0 + \mathbf{B}\mathbf{u}'(t))\mathbf{P}, \tag{6}$$

where $\mathbf{P} \in \mathbb{R}_{\geq 0}^{n^{N_c N'}}$ is the probability mass vector for all possible states; $\mathbf{A}_0 \in \mathbb{R}^{n^{N_c N'} \times n^{N_c N'}}$

is an uncoupled infinitesimal generator; $\mathbf{u}'(t) \in \mathbb{R}_{\geq 0}^{n^{N_c N'}}$ is a (potentially time varying) vector of control inputs with one entry for every distinct state in the system's state space; and $\mathbf{B}$ is a fixed tensor that operates on $\mathbf{u}'$ to adjust the master equation to account for the optogenetic input. Explicit examples for the construction of quantities $\mathbf{A}$ and $\mathbf{B}\mathbf{u}'$ for different control laws are provided in Methods.

At first, we consider the special case where the control signal depends only on the current state at each instant in time. In this case, the vector $\mathbf{u}'$ is constant with respect to time and depends only on the enumeration of the possible states. As such, the infinitesimal generator in Eq. (6) reduces to a time-homogeneous master equation for a standard discrete state Markov process. We note that the control signal $\tilde{u}'(t)$ may still fluctuate due to changes in $\tilde{\mathbf{X}}$ and can be written using the notation $\tilde{u}'(t) = \mathbf{u}_{\tilde{\mathbf{X}}'(t)}$ to denote that the index of the vector $\mathbf{u}'$ is specified by the instantaneous state $\tilde{\mathbf{X}}(t)$.

By changing the specification of $\mathbf{u}'$, we can further simplify this class to consider several different types of state-based control rules. Here we consider three different possibilities: an UnAware Controller (UAC) that has no information about any cells to make its control decision:

$$\mathbf{u}^{\text{UAC}}_{\tilde{\mathbf{X}}(t)} = u^{\text{UAC}} = \text{constant, where } \mathbf{u}^{\text{FAC}} \text{ is the same for every } \tilde{\mathbf{X}}(t)); \tag{7}$$

a Fully Aware Controller (FAC) which has complete knowledge of all cells $\tilde{x}_1$, $\tilde{x}_2$, ... and therefore uses the full state vector to make its control decisions:

$$\mathbf{u}^{\text{FAC}}_{\tilde{\mathbf{X}}(t)} = u^{\text{FAC}}(\tilde{x}_1, \tilde{x}_2, \ldots), \text{ where } \mathbf{u}^{\text{FAC}} \text{ has a unique element for every } \tilde{\mathbf{X}}(t), \tag{8}$$

and a Partially Aware Controller (PAC) which uses only information of a single cell, e.g., x_1, to make its control decisions:

$$\mathbf{u}^{\text{PAC}}_{\tilde{\mathbf{X}}(t)} = u^{\text{PAC}}(\tilde{x}_1), \text{ where } \mathbf{u}^{\text{PAC}} \text{ changes only with } \tilde{x}_1(t). \tag{9}$$

For our specific case of model $\mathcal{M}'_A$, the FSP truncation size is $n = 50$, and the number

of species is $N' = 1$; thus for two cells ($N_c = 2$), the matrix $\mathbf{A}_0 \in \mathbb{R}^{2500\times 2500}$.

With the formal definition of the cost function J from Eq. (5) and the CME from Eq. (6), we can then compute the gradient of the cost function with respect to the control law as (see derivation in SI):

$$\nabla_{\mathbf{u}}(J) = \mathbf{C}(\mathbf{A}_0 + \mathbf{B}\mathbf{u}')^{-1}\mathbf{B}\mathbf{P}. \tag{10}$$

Having specified the cost function gradient with respect to the controller, we run an optimization algorithm (see Methods) to search along the gradient to find local minima for the cost function. We note again that for each of these optogenetic Maxwell's Demon control strategies, every cell experiences equal inputs at every instant in time, although the magnitude of that single input changes in time as the cells fluctuate to different states.

Starting with the unaware (UAC) and fully aware (FAC) controllers, we used the FSP approach and local optimization (see SI for details on the optimization procedure) to find a locally optimal control strategy in the form of the constant u^{UAC} or the two-dimensional scalar field $u^{FAC}(x_1, x_2)$, and with and without the addition of auto-regulation to the model. The resulting optimal controllers are presented in the top row of Fig. 4, where the color at each point represents the control input magnitude $u^{xAC}(x_1, x_2)$ as a function of the species quantities x_1 for the first cell and x_2 for the second cell. In the figure, a vector field of white arrows depicts the net direction of probability flow due to the combined action of the internal auto-regulatory effects and the feedback control input $u^{xAC}(x_1, x_2)$ on the system. For practical implementation, the control input $u'(\phi) = u^{xAC}(x_1, x_2)$ would be converted to light intensity through inversion of the calibration curve in Fig. 2B (blue line). The middle row of Fig. 4 shows the resulting steady state joint distribution of each condition, and in the bottom row of Fig. 4 shows the corresponding marginal distributions, with cell one in solid blue lines and cell two in dashed red lines. Fig. 4 shows that symmetry is broken only in the case where the cells' genetic design includes auto-regulation *and* the feedback

controller contains knowledge of the cells (i.e., the far right row of Fig. 4). In this best-case scenario, the two cells are very effectively driven each to their own unique and pre-chosen phenotype, irrespective of their initial conditions. If feedback is included without auto-regulation, the cells' distributions are made tighter at some intermediate value between the target values for cell one and cell two; this results in a slightly better numerical cost value, but it becomes even less likely that both cells will reach their target phenotypes at the same time as compared to the uncontrolled situation (compare first and second row in Fig. 4). Conversely, if auto-regulation is included without feedback (i.e., the light level is fixed at some optimal value), the cells exhibit a bimodal distribution with some cells near each target value, but there is no means to control which cell expresses which phenotype, and the cost function is again worse than the case with no auto-regulation (compare first and third rows of Fig. 4). These data taken together suggest that auto-regulation and feedback control, in addition to intrinsic single-cell noise, are all critical to the break symmetry and enable differential control of multiples cells using a single input. Specifically, noise breaks the symmetry of cell behavior and allows cells to switch independently between phenotypes, feedback helps to reinforce this noise and steer cells toward desired phenotypes, and auto-regulation helps stabilize cell behaviors once they have attained their desired phenotypes.

Effective differential control of many cells using a single input is possible, even when observations are limited to a single cell of interest.

We next examined a more general problem to control an arbitrary number of cells simultaneously. In this case, we consider a situation where the controller acts on many cells simultaneously with the goal of steering a single observed cell (x_1) to one state and all remaining unobserved cells to another state. To handle this problem, we first utilize a partially aware controller (PAC) that observes only $\tilde{x}_1(t)$ and ignores the states of all other cells. Fig. 5 compares the results of this simplified controller (middle column) to those of the two-cell controller from the previous section (left column). The resulting control law, $u^{PAC}(x_1)$, is optimized to find the best control signal input for

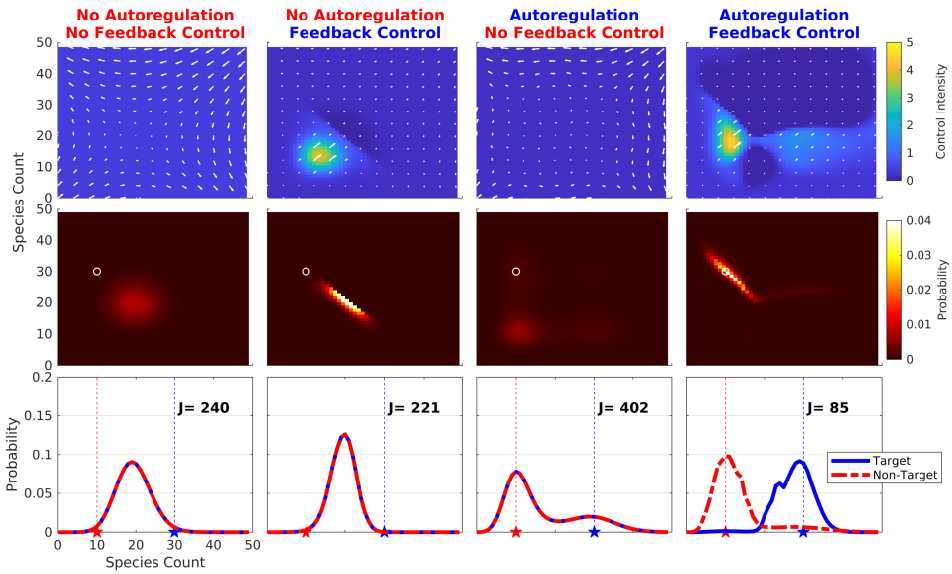

Figure 4: Optimized control laws and performance results for cells and different combinations of regulatory structure and control strategy. (Top row) Optimized control input versus cells' states, $\mathbf{u}^{xAC}(x_1, x_2)$ (colormap at far right). White vectors plotted over the control inputs represent the net flow of probability at that point in state space. (Middle row) Resulting joint probability distribution ($\mathbf{P}(x_1, x_2)$, colormap at far right). The target point $T = [30, 10]^T$ is denoted by a white circle. (Bottom row) Corresponding marginal distributions for the cells $\mathbf{P}(x_1)$ (blue) and $\mathbf{P}(x_2)$ (red) and time-averaged cost function, J. Leftmost column shows results with no auto-regulation and a constant control signal. Second column shows the result with no auto-regulation and feedback control strategy. Third column shows the result with auto-regulation and constant control signal (UAC in text). Rightmost column shows the results for auto-regulation and fully aware feedback control (FAC in text).

each possible value x_1 of the observed cell, and the top row of Fig. 5 shows that this optimal PAC controller depends only on the observed cell (x_1 axis) but is constant with respect to all unobserved cells (x_2 axis). Despite the simplicity of this control strategy and the fact that it requires only knowledge of the instantaneous expression of the single observed cell, the second and third rows of Fig. 5 show that the PAC controller effectively breaks symmetry to force the observed cell to a high expression phenotype while most unobserved cells are correctly directed to the low expression phenotypes.

Although the partially aware controller under-performs compared to the fully aware controller for the case of exactly two cells (compare middle and left columns of Fig. 5), the advantage of the PAC is that it works equally well, and without any modification, for any arbitrarily large number of unobserved cells (Supplemental Fig. S1). In contrast, to use the FAC for more than two cells requires modification, such as training of a higher rank tensor representation of the control algorithm, or defining a control law based on the mean, median, or some other statistical quantity for the groups of cells to be assigned to each phenotype. Unfortunately, the former high-order tensor approach is computationally intractable using existing methods, and the latter approach rapidly loses performance as the number of cells is increased. For example, when the controller is based on the observation of $\tilde{x}_1$ and the mean of the remaining cells $\{\tilde{x}_2, \tilde{x}_3, \ldots\}$ (FACM, see Supplemental Information), we observe that for any more than a single cell in the second group, the PAC outperforms the FACM (Supplemental Fig. S1, compare FACM and PAC controllers).

A probabilistic Model Predictive Controller (pMPC) can improve the control of many cells using a single observer and a single input signal.

In the previous section, the PAC control was based on only the observation of a single observed cell, and had no information about the other cells that it was also seeking to control. However, knowing the history of the input signal (i.e., the light intensity over time in the past), the FSP approach allows for the possibility that the controller can estimate the probability distribution of all non-observed cells. With this possibility

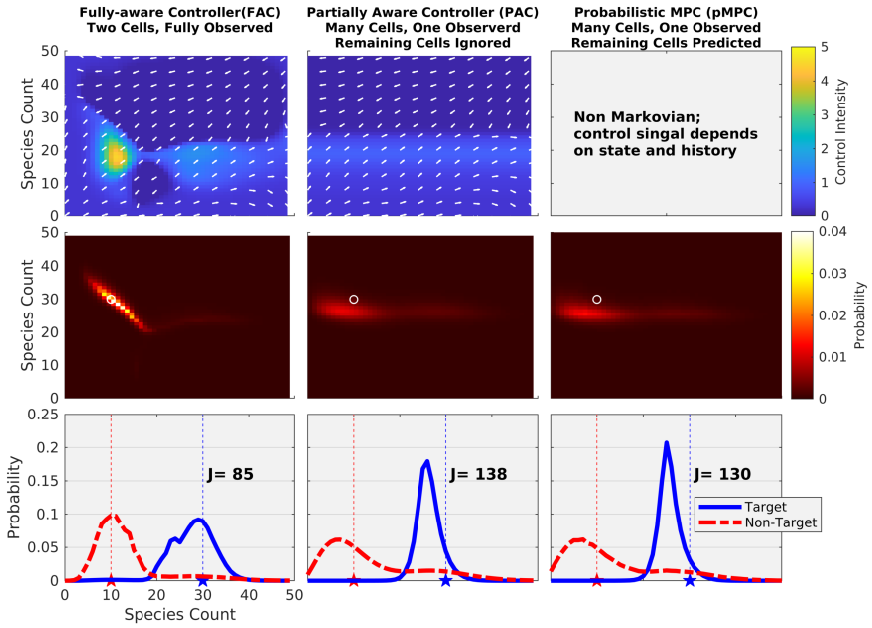

Figure 5: Control laws and performance for strategies to control many cells at once using partial knowledge only of a single observed cell. Each row shows: (top) the control law $\mathbf{u}'_{\tilde{\mathbf{X}}} = u^{xAC}(\tilde{x}_1, \tilde{x}_2)$, (middle) the resulting joint probability distribution ($\mathbf{P}(x_1, x_2)$), and (bottom) the marginal distributions for the observed cell $\mathbf{P}(x_1)$ (blue) and remaining cells $\mathbf{P}(x_2)$ (red). The middle column shows results for the partially aware controller (PAC in text), where the control depends only on the single observed cell ($\mathbf{u}'_{\tilde{\mathbf{X}}} = u^{PAC}(\tilde{x}_1)$). The rightmost row shows results for the probabilistic model predictive controller (pMPC), which also applies to an arbitrary number of cells. For the pMPC, the process is non-Markovian in that the controller ($\mathbf{u}'_{\tilde{\mathbf{X}}} = u^{\mathrm{pMPC}}(\tilde{x}_1, \tilde{\mathbf{P}}(x_{i\neq 1}))$ depends not only upon the state of observed $\tilde{x}_1$, but also on the predicted probability vector for the unobserved cells, $\tilde{\mathbf{P}}(x_{i\neq 1})$. To enable comparison to previous two-cell cases, the leftmost column shows the results for the fully aware feedback control (FAC) with only two cells. Color bars are shown to the right of each row.

in mind, we next explored a new class of controller that could use direct knowledge of the protein expression in the observed cell, the known control input signal at the current time, and a probabilistic model to predict the distribution of expression in the unmeasured cells. Specifically, we use the FSP approach to integrate our prediction of the probability distribution for the unobserved cells as:

$$\frac{d}{dt}\tilde{\mathbf{P}}_{uo} = \left(\mathbf{A}_0 + \mathbf{B}\tilde{u}'(t)\right)\tilde{\mathbf{P}}_{uo}, \tag{11}$$

where $\tilde{\mathbf{P}}_{uo} \in \mathbb{R}^n_{\geq 0}$ is the estimate of the probability mass vector for the protein expression in the unobserved cells, $\mathbf{A}_0 \in \mathbb{R}^{n \times n}$ is the infinitesimal generator for a single cell in the absence of any control input, and the scalar variable $\tilde{u}'(t) > 0$ is the instantaneous input signal that is produced by the controller. We note that the probability mass vector estimate $\tilde{\mathbf{P}}_{uo}$ is the result of a stochastic process that depends upon the full history of the input signal $\tilde{u}'(t)$.

Using this prediction for the unobserved cells, the pMPC controller law can now be defined as

$$u^{\mathrm{pMPC}}(\tilde{x}_1, t) = c_{\tilde{x}_1} + \mathbf{z}_{\tilde{x}_1} \tilde{\mathbf{P}}_{uo}(t) \tag{12}$$

or written in vector form for all possible values of $\tilde{x}_1$ as:

$$\tilde{\mathbf{u}}^{\mathrm{pMPC}}(t) = \mathbf{c} + \mathbf{Z}\tilde{\mathbf{P}}_{uo}(t) \tag{13}$$

where $\mathbf{c} = [c_0, c_1, \ldots, c_{n-1}]^T$ is a constant vector in $\mathbb{R}^n$, and $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is a matrix of linear weights which adjusts the input based off of the estimated unobserved probability distribution $\tilde{\mathbf{P}}_{uo}$. In our practical implementation, we assume that the controller in Eq. (13) is piecewise constant with respect to $\tilde{\mathbf{P}}_{uo}$ over a time step of 0.5 min, but it changes instantaneously with each even that affects the observed cell $\tilde{x}_1$. The weights of $\mathbf{c}$ and $\mathbf{Z}$ are then jointly optimized to minimize the cost function, J. The simple formulation of the control law in Eq. (13) admits the possibility for non-achievable negative values of light in order to construct a computationally tractable optimization procedure. However, in testing the controller, this non-physical situation is corrected by saturating negative control signals to zero in the true test of the system (see Supplemental Information). We note that this approximation to allow for negative control signals in the control law specification and subsequent correction to saturate these to zero in the control law test suggest that the pMPC controller identified here is sub-optimal. However, despite this non-optimal design, Fig. 5 shows that the resulting *non-optimized* pMPC (J=130) controller outperforms the *fully optimized* PAC (J=138), demonstrating that probabilistic model predictions can be used to improve control performance even in

the absence of observations for many of the cells under its control. Having succeeded in our main goal to determine if probabilistic predictions could improve control results in principle, we leave the fine tuning of the specific pMPC control strategy to future investigations and more sophisticated control design strategies.

For a closer look at how the pMPC approach works to control observed and unobserved cells alike, Figs. 6A shows an example control input over time, and Fig. 6C shows the resulting trajectories over time for the observed cell (blue), the predicted probability distribution for unobserved cells (gray shading), and a representative unobserved cell (red). We reiterate that the controller has no direct knowledge of the red line. From the figure, it is clear that observed cell is well maintained near to its target value with low variability. Moreover, Fig. 6C shows that knowledge of the fluctuating input signal is sufficient to yield good predictions of the unobserved cell response (compare red line with dark gray shading), although as expected there are periods of poor predictions when the specific unobserved cell samples the higher or lower tail of the predicted distribution (e.g., at about 1900 minutes for the red curve in Fig. 6C). In addition to outperforming the PAC approach in terms of the overall cost function, the pMPC provides additional predictions for when the controller is effective, or when unobserved cells are more likely to escape from their intended phenotype. To illustrate this Figs. 6B shows the probability of observing fifteen or less protein molecules in the unobserved cell. When this probability exceeds a 90 percent threshold the control is expected to be effective, and the region is labeled as orange in Figs. 6B and Figs. 6C. Figs. 6D shows the marginal distributions averaged over all times, and Fig. 6E shows the marginal distribution only for the periods of time when the probabilistic model predicts its own effective control of the unobserved cells (orange regions). By focusing only on these times identified as successful by the controller, the cost function of the controller substantially decreases from $J = 130$ to $J = 58$. These results suggest that predicted dynamic information about the unobserved cells can not only be used to improve the quality of the controller, but that the pMPC can also be used to self-assess when control is working well, and when it is not.

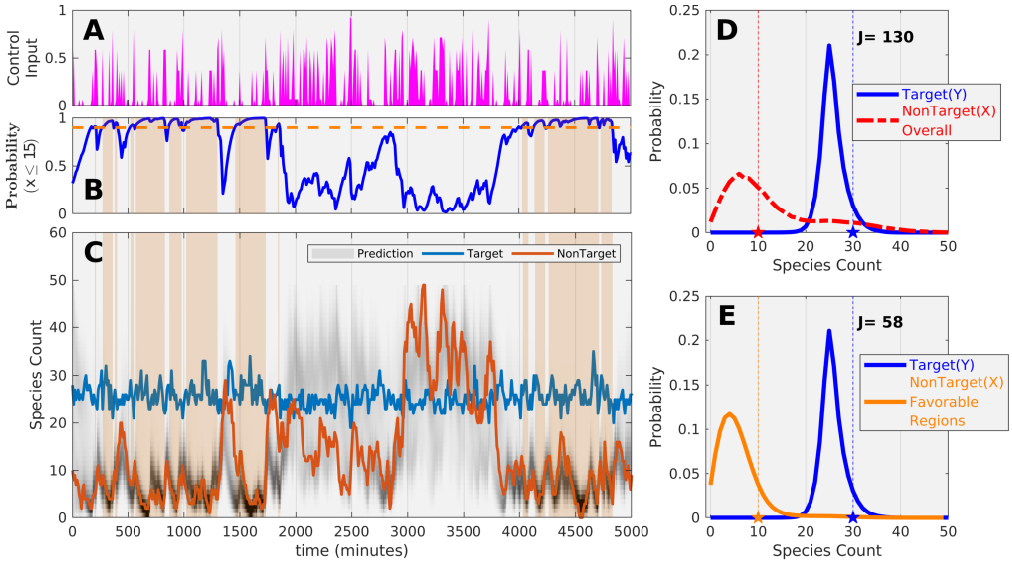

Figure 6: Control signal and response versus time for pMPC control. (A) Control signal generated by the pMPC controller. (B) Probability that unobserved cell is 15 or less versus time. When the probability is greater than 0.9 (orange line), then the system is considered to be in a "effective control" state. (C) Predicted transient distributions for unobserved cells (gray), observed cell (blue), and a single unobserved cell (red). Orange regions correspond to effective control times. (D) Time averaged performance of the control law in terms of marginal distributions for the observed call (blue) and unobserved cells (red). (E) Average of the pMPC performance, when considering only effective control periods.

Controllers designed using simplified models can be effective to control more complicated processes with hidden mechanisms and dynamics.

We next ask how well could controllers designed using simplified stochastic models work when they are applied to control more complex systems that contain additional hidden states and which have unknown dynamics or time delays. To perform this analysis, we first account for the difference in meaning and units for the input signal $u'(\phi)$ used in the reduced auto-regulation models ($\mathcal{M}'_A$ in Fig. 1C) and its analog $u(\phi)$ used in the full auto-regulation model ($\mathcal{M}_A$ in Fig. 1D). By using steady state ODE analyses of $\mathcal{M}'_A$ and $\mathcal{M}_A$, we quantified the calibration curve to map inputs between the two models as shown in Fig. 7A. After calibration of the input signals, we verified that the full and reduced auto-regulation models result in similar bifurcation diagrams as shown in Fig. 7B. However, although calibration allows us to match both the quasi-steady (i.e., very slow) and fast fluctuating input responses of the two models (Figs. 7C,E, respectively), the temporal responses to slow input frequencies are qualitative and quantitatively different, as can be observed by the different input-to output time lags and amplitudes in Fig. 7D.

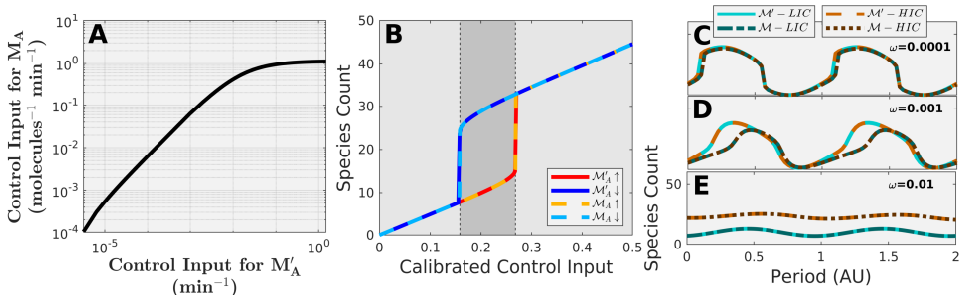

Figure 7: Calibration and use of controllers for use with a new, more complex model. (A) Calibration curve identified to match steady state ODE of simple and complex auto-regulation model ($\mathcal{M}'_A$ and $\mathcal{M}_A$ from Fig. 1C,D). (B) Steady state analyses show that the calibrated inputs result in similar hysteresis behavior for both $\mathcal{M}'_A$ and $\mathcal{M}_A$. (C) Input-output response analysis shows that $\mathcal{M}_A$ with calibrated inputs closely matches behavior of $\mathcal{M}'_A$ at ultra slow frequencies (0.0001 RPM), but (D) the more complex model begins to lag at slow frequencies (0.001 RPM). (E) At fast frequencies of 0.01 RPM, the complex auto-regulation model $\mathcal{M}_A$ is able to retain memory of its initial conditions and again exhibits similar phenomena compared to the simplified model.

Having calibrated the controller for the full model to match the response of the simpler model, we then take the UAC, FAC, and PAC controllers from above and apply them directly (i.e., without any further tuning or optimization) to the full mechanistic model with auto-regulation. Despite the differences in temporal behaviors between the two models, the previously identified UAC and FAC controllers still work to break symmetry and drive both cells toward the correct differentiated phenotypes as shown in Fig. 8. We note that with further modifications, the control laws derived using the simplified model could certainly be improved for use in the more complex system. However, our primary goal was to explore how well designs made in one context should perform when used in another different context, and subsequent fine tuning for the complex model is left for future investigations.

Conclusion

The treatment of noise in synthetic biology has largely been centered around the management of noise as a nuisance property that needs to be mitigated or eliminated. Despite improvements to minimizing noise in bio-circuits, noise largely remains a fundamental physical limit due to the combination of very small cell sizes, where single molecular events have increased importance, and increasing complexity of synthetic circuits, where most dynamical influences are unknown or unmeasured. The results here show how a few increasingly common synthetic biology motifs–such as optogenetic transcription factors, activatable polymerases and auto-regulation promoters–can in principle be combined to form regulatory modules and integrated with new external feedback controllers not only to mitigate intrinsic noise, but even to exploit that noise to achieve new multi-cellular behaviors.

Our work compliments that of previous efforts to develop single- and multiple-input-single-output (SISO and MISO) control for synthetic biology applications[40,41] that have primarily sought to control one cell at a time or to control an entire population of cells to all reach the same phenotype. Specifically, we have demonstrated new

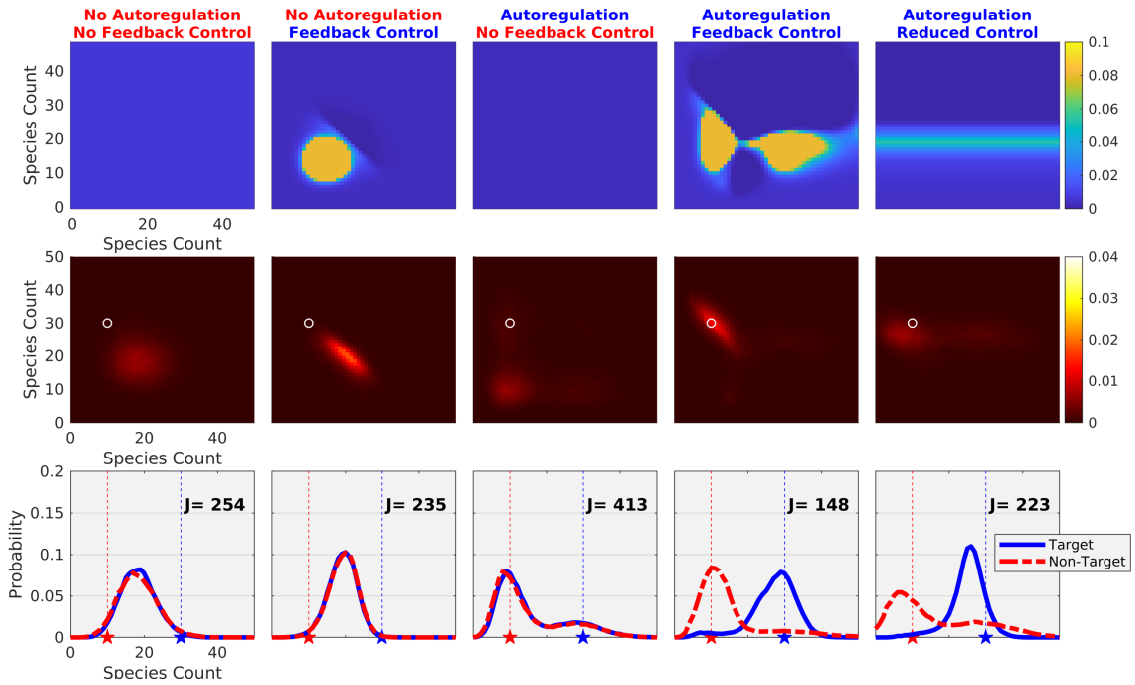

Figure 8: Full Models are paired with the calibrated controllers (top row) to solve for the joint probability distributions (middle row) and marginal probability distributions (bottom row). As before both auto-regulation and feedback are needed to break symmetry and control fails if either of these are missing (rightmost three columns). The fully aware controller (FAC, fourth column) successfully works to control two cells with the complex dynamics, and the partially aware controller (PAC, rightmost column) successfully can control a single observed cell to one phenotype and an arbitrary number of unobserved cells to another different phenotype.

types of optimizable single-input-multiple-output (SIMO) stochastic controllers that rely on the integration of noise, non-linear auto-regulation, and feedback to simultaneously control of multiple cells using a single chemical or optogenetic input. The first of these, the fully aware controller (FAC), assumes full knowledge of each individual cell's behavior and achieves the best control performance. The disadvantage of the FAC is that it requires knowledge of each individual cell (i.e., optical tracking and image processing analyses) and computationally intensive operations both to solve multi-dimensional chemical master equations and to search very high dimensional spaces for optimal controllers. However, a second partially aware controller (PAC) requires only the knowledge of a single cell of interest, yet the PAC can control that cell to one phenotype and drive all others to an alternate phenotype with an accuracy almost equal to the FAC. The advantage of the PAC is that it is very easy to implement and optimize as the dimension of the control law must only consider the single observed cell. The third controller introduces a probabilistic model predictive control (pMPC) strategy that computationally predicts the probability distribution of all non-observed cells based on integration of the chemical master equation under the known history of the applied input signal. Although we envision that similar control strategies may have applications in other fields, such as for autonomous vehicle or smart grid applications, to our knowledge, the proposed pMPC approach is the first example of a hybrid control strategy that predicts and exploits noise and feedback to simultaneously and differentially control multiple identical agents using a single control signal.

In addition to noise, one of the most important challenges in model-driven synthetic biology, is that many important regulatory mechanisms are currently unknown, even for simple biological systems. Similarly, very few parameters are known at any level of certainty, and most of these parameters vary from cell to cell or situation to situation. Moreover, it is already extremely computationally expensive to combine all known mechanisms into a single computational model, and although such models can be useful to reproduce a variety of biological behaviors,[51,52] such whole cell models are far too inefficient to enable the vast numbers of different simulations needed to optimize a

design or control strategy. To circumvent these concerns, we demonstrated how a highly simplified phenomenological model could be used to design a controller that could be easily recalibrated using steady state dose-response measurements and then applied directly to a control a more complex system with hidden dynamics and with qualitatively and quantitatively different dynamic response features. Although it is common practice to use simple deterministic models to guide engineering design of modern complex devices, this demonstration in the context of stochastic single-cell processes suggests that there is also hope for similar applications of simple models in synthetic biology.

Although the potential of our computational results remains to be verified through independent experimental investigation, we believe that this numerical demonstration of the potential for a new control paradigm not only opens new possibilities for integrated "cyber organic" approaches in synthetic biology,[28–30,45] but could also offer insight into natural cellular differentiation processes where cellular states are sensed, and control signals are transmitted, by neighboring cells. For example, it has been suggested that stochastic fluctuations in expression lead embryonic stem cells to achieve substantial, and functionally relevant heterogeneity in Nanog expression, where transiently low Nanog expression cells are prone toward differentiation, whereas high Nanog expression cells are less likely to differentiate.[53] As such, it might be interesting to explore the possibility that temporally controlled fluctuations in Nanog transcription factors[54,55] could selectively direct specific neighboring cells to differentiate while maintaining others in the stem state. Overall, we envision that advancing synthetic biology motifs, especially an increasing diversity of orthogonal transcription factors and promoters,[12,56] improved live cell reporters,[13,57] and faster and more specific optogenetically controlled transcription factors inputs,[30] will integrate synergistically with new probabilistic model predictive control analyses to improve future efforts to understand how noise, non-linearity, and feedback combine to drive cell fate decisions in applications ranging from synthetic biofuel and biomaterial production to developmental dynamics or regenerative medicine.

Methods

Definition of Models in Terms of Stoichiometries and Propensities

To introduce our numerical approaches, consider a generic cell regulatory process that contains N' distinct chemical species that interact with each other through M' different reactions. At any point in time, the current state of the process in a *single* cell can be described by an N'-element vector $\mathbf{x} = [x_1, \ldots, x_{N'}]^T \in \mathcal{X}$, where $\mathcal{X}$ denotes the set of all possible states (e.g., the nonnegative spaces $\mathbb{R}^{N'}_{\geq 0}$ for a continuous process or $\mathbb{Z}^{N'}_{\geq 0}$ for a discrete process). The definition of the full state $\mathbf{X}$ for *multiple* cells is easily concatenated to consider a set of N_c individual cells

$$\mathbf{X} = [x_1^{\text{cell } 1}, \ldots, x_{N'}^{\text{cell } 1}, \ldots, x_1^{\text{cell } N_c}, \ldots, x_{N'}^{\text{cell } N_c}]^T,$$

with an appropriate change to the total numbers of species ($N \equiv N_c N'$) and reactions ($M \equiv N_c M'$).

Under the assumption of a well-mixed spatial environment within each cell, one can define the dynamics of such a process by specifying the reaction stoichiometry vector and reaction rate for each μth reaction.[58] The *stoichiometry vector*, $\mathbf{s}_\mu \in \mathbb{Z}^N$ is the net integer change in molecules after exactly one event of the μth chemical reaction (i.e., $\mathbf{s}_\mu \equiv \mathbf{X}(\text{after } \mu^{\text{th}} \text{ reaction}) - \mathbf{X}(\text{before } \mu^{\text{th}} \text{ reaction})$). For continuous processes, the *reaction rate*, $f_\mu(\mathbf{X}, \mathbf{\Lambda}, u)$ is a scalar that defines the speed at which the μth reaction would be expected to occur given the current state $\mathbf{X}(t)$, fixed physical parameters $\mathbf{\Lambda}$ and time- or state-varying control parameter $u(\mathbf{X}, t)$. For discrete stochastic chemical reactions, the reaction rate is replaced with a *propensity function* $w_\mu(\mathbf{X}, \mathbf{\Lambda}, u)dt$, which describes the probability that a single μ^{th} reaction would occur in the next infinitesimal time step of length dt given $\mathbf{X}(t)$, $\mathbf{\Lambda}$, and $u(\mathbf{X}, t)$. For reduced order models $\mathcal{M}'_U$ and $\mathcal{M}'_A$, we replace u with u' to denote the change in units needed for consistency with the model reduction.

ODE Representation of Models

Using these simple definitions, one can easily write an ordinary differential equation (ODE) to define a deterministic description of the process dynamics as:

$$\frac{d\mathbf{X}}{dt} = \sum_{\mu=1}^{M} \mathbf{s}_\mu f_\mu(\mathbf{X}, \mathbf{\Lambda}, \mathbf{u}), \tag{14}$$

$$= \mathbf{S}\mathbf{f}(\mathbf{X}, \mathbf{\Lambda}, u(\mathbf{X}, t)), \tag{15}$$

where $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_M] \in \mathbb{Z}^{N\times M}$ is the stoichiometry matrix, and $\mathbf{f}(\mathbf{X}, \mathbf{\Lambda}, u(\mathbf{X}, t)) = [f_1, \ldots, f_M]^T \in \mathbb{R}^M > 0^M$ is the vector of reaction rates. For any given stoichiometry matrix $\mathbf{S}$, and reaction rate function vector, $\mathbf{f}(\mathbf{X}, \Lambda, u)$, the rate of change of $\mathbf{X}$ described by Eq. (14) can be integrated numerically to describe the system dynamics over time.

Discrete Stochastic Representation of Models

For discrete stochastic systems, the specification of the reaction stoichiometry and propensity functions is sufficient to generate individual trajectories of the process using Gillespie's Stochastic Simulation Algorithm (SSA,[47]). Alternatively, one can also use these two properties to uniquely define the Chemical Master Equation (CME,[59]) as:

$$\frac{d}{dt}P(\mathbf{X}) = -\sum_{\mu=1}^{M} w_\mu(\mathbf{X}(t), \mathbf{\Lambda}, u(\mathbf{X}, t))P(\mathbf{X}) + \sum_{\mu=1}^{M} w_\mu(\mathbf{X} - \mathbf{s}_\mu, \mathbf{\Lambda}, u(\mathbf{X} - \mathbf{s}, t))P(\mathbf{X} - \mathbf{s}_\mu). \tag{16}$$

For this discrete state description, one can always enumerate all possible states as $\{\mathbf{X}_1, \mathbf{X}_2, \ldots\} \equiv \mathcal{X}$ and define a probability mass vector as the similarly ordered probabilities, $\mathbf{P} \equiv [P(\mathbf{X}_1), P(\mathbf{X}_2), \ldots]^T$. Because the CME in Eq. (16) is linear in every term $P(\mathbf{X}_i)$, it is often written in matrix format:

$$\frac{d}{dt}\mathbf{P} = \mathbf{A}\mathbf{P}, \tag{17}$$

where the square matrix $\mathbf{A}$ is known as the *infinitesimal generator* and is defined directly from Eq. (16) as

$$A_{ij} = \begin{cases} -\sum_{\mu=1}^{M} w_\mu(\mathbf{X}_j, \mathbf{\Lambda}, u(\mathbf{X}_j, t)), & \text{for } i = j, \\ w_\mu(\mathbf{X}_j, \mathbf{\Lambda}, u(\mathbf{X}_j, t)), & \text{for } \mathbf{X}_i = \mathbf{X}_j + \mathbf{s}_\mu \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

We note that the summation of $\mu = 1$ to $M = N_c M'$ accounts for the increase in the number of possible reactions due to the existence of multiple cells. Because each term $\mathbf{X}_i$ refers to a specific enumerated state vector that is fixed in time, in the special case where $u \equiv u(\mathbf{X}_i)$ (i.e., where u depends only on the current state and does not depend *explicitly* on time), the matrix $\mathbf{A}$ is constant with respect to time. For convenience, we can define the control parameter in vector form $\mathbf{u} = [u(\mathbf{X}_1), u(\mathbf{X}_2), \ldots]$. The final CME model with control can be written in simple form by separating the infinitesimal generator into it basal and control induced components as:

$$\frac{d}{dt}\mathbf{P} = (\mathbf{A}_0 + \mathbf{B}\mathbf{u})\mathbf{P}. \tag{19}$$

In this formulation, although the dynamics of each identical cell was independent and uncoupled in the basal infinitesimal generator $\mathbf{A}_0$, the added infinitesimal generator from the control input, $\mathbf{B}\mathbf{u}$, can introduce coupling between cells. As an example, consider the fully aware controller (FAC) for two cells. The ith state is written $\mathbf{X}_i = [x_{i1}, x_{i2}]^T$, and the control infinitesimal generator can be written as:

$$[\mathbf{B}\mathbf{u}]_{ij}^{FAC} = \begin{cases} -2u^{FAC}(x_{i1}, x_{i2}), & \text{for } i = j, \\ u^{FAC}(x_{i1}, x_{i2}), & \text{for } \mathbf{X}_i = \mathbf{X}_j + [1, 0]^T, \\ u^{FAC}(x_{i1}, x_{i2}), & \text{for } \mathbf{X}_i = \mathbf{X}_j + [0, 1]^T, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Similarly, for the partially aware controller (FAC) for two cells, the control infinitesimal generator can be written as:

$$[\mathbf{Bu}]_{ij}^{PAC} = \begin{cases} -2u^{PAC}(x_{i1}), & \text{for } i = j, \\ u^{PAC}(x_{i1}), & \text{for } \mathbf{X}_i = \mathbf{X}_j + [1,0]^T, \\ u^{PAC}(x_{i1}), & \text{for } \mathbf{X}_i = \mathbf{X}_j + [0,1]^T, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

As discussed in the main text, in either case, the coupling introduced by the control infinitesimal generator $[\mathbf{Bu}]^{FAC}$ or $[\mathbf{Bu}]^{PAC}$ is sufficient to break symmetry and encourage cells toward desired differential expression phenotypes.

Solution Scheme for Chemical Master Equation

To solve the CME in Eq. (19), we use the Finite State Projection (FSP[60]) approach, which truncates the allowable state space for every species and results in a finite dimensional ODE. However, it should be noted that the state space of a single arbitrary chemical species is given by the ordered set $[0, 1, ..., n-1]$ up to some truncation limit n. The state space of multiple species, is enumerated by forming a tuple of all possible species available, each up to a similar maximum number. For a system of N_c cells with N' chemically reacting species, where each species can range up to a maximum of $n-1$ copies per cell, the number of distinct states is $n^{N'N_c}$, which quickly becomes intractable when n, N_c, or N' is large. For this reason, model reductions, simplifications, or approximations are essential, especially when these models are to be used with millions of different parameter sets when searching for optimal control strategies. Further, it is important to test and verify if control strategies designed and optimized using such simplified models will continue to be effective when applied to more general and more complex systems.

Fitting of models to data

Fits of the reduced model to experimental data was performed numerically by optimizing both the set of model parameters and the calibration variables in unison. Since the parameter fits of $\mathcal{M}'_U$ did not reveal a single set of unique parameters which fit data, the decay rate was calculated by hand by fitting the middle region of Fig. 2A and then fitting the parameters after fixing the protein decay rate. Fitting the $\mathcal{M}_U$ to experimental data using calibration was performed by hand since mathematical tools to fit data often yielded poor results by becoming stuck in local minima. These hand fits were also constrained such that the decay rate of the protein is the same decay rate in $\mathcal{M}'_U$.

Supporting information

Supporting information is one file containing:

Supporting Text - supplemental details on feedback control designs.

Supporting Figure S1 - Visualization of pMPC Control Law

Supporting Figure S2 - Performance of stochastic controllers using varying numbers of non-target cells.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM124747. The work reported here was partially supported by a National Science Foundation grant (DGE-1450032). Any opinions, findings, conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contribution

BM designed and acquired funding for the study. MPM and BM developed computational and theoretical tools. MPM performed all computational analyses. MPM wrote the manuscript. MPM and BM edited the manuscript.

Conflicts of Interest

None

Data Availability

All data and computational codes needed to reproduce the figures in this manuscript are freely available through the GitHub page: `https://github.com/MunskyGroup/Michael_May_et_al_2021`

References

(1) Hasty, J.; McMillen, D.; Collins, J. J. Engineered gene circuits. *Nature* **2002**, *420*, 224–230

(2) Knight, T. Idempotent Vector Design for Standard Assembly of Biobricks. *MIT Libraries* **2003**, 1–11

(3) Jang, S. H.; Cha, J. W.; Han, N. S.; Jeong, K. J. Development of bicistronic expression system for the enhanced and reliable production of recombinant proteins in Leuconostoc citreum. *Scientific Reports* **2018**, *8*, 1–11

(4) Menolascina, F.; Fiore, G.; Orabona, E.; De Stefano, L.; Ferry, M.; Hasty, J.; di Bernardo, M.; di Bernardo, D. In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks. *PLOS Computational Biology* **2014**, *10*, 5

(5) Gardner, T.; Cantor, C.; Collins, J. Construction of a Genetic Toggle Switch in Escherichia coli. *Nature* **2000**, *403*, 339–342

(6) Wu, M.; Su, R. Q.; Li, X.; Ellis, T.; Lai, Y. C.; Wang, X. Engineering of regulated stochastic cell fate determination. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10610–10615

(7) Kramer, B. P.; Viretta, A. U.; Daoud-El Baba, M.; Aubel, D.; Weber, W.; Fussenegger, M. An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **2004**, *22*, 867–870

(8) Schaumberg, K. A.; Antunes, M. S.; Kassaw, T. K.; Xu, W.; Zalewski, C. S.; Medford, J. I.; Prasad, A. Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nature Methods* **2016**, *13*, 94–100

(9) Shin, J.; Zhang, S.; Der, B. S.; Nielsen, A. A.; Voigt, C. A. Programming Escherichia coli to function as a digital display. *Molecular Systems Biology* **2020**, *16*, 3

(10) Ng, A. H.; Nguyen, T. H.; Gómez-Schiavon, M.; Dods, G.; Langan, R. A.; Boyken, S. E.; Samson, J. A.; Waldburger, L. M.; Dueber, J. E.; Baker, D.; Others, Modular and tunable biological feedback control using a de novo protein switch. *Nature* **2019**, *572*, 265–269

(11) Moon, T. S.; Lou, C.; Tamsir, A.; Stanton, B. C.; Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **2012**, *491*, 249–253

(12) Groseclose, T. M.; Rondon, R. E.; Herde, Z. D.; Aldrete, C. A.; Wilson, C. J. Engineered systems of inducible anti-repressors for the next generation of biological programming. *Nature Communications* **2020**, *11*, 1

(13) Morisaki, T.; Lyon, K.; DeLuca, K. F.; DeLuca, J. G.; English, B. P.; Zhang, Z.; Lavis, L. D.; Grimm, J. B.; Viswanathan, S.; Looger, L. L.; Lionnet, T.; Stase-

vich, T. J. Real-time quantification of single RNA translation dynamics in living cells. *Science* **2016**, *352*, 1425–1429

(14) Liu, C. C.; Jewett, M. C.; Chin, J. W.; Voigt, C. A. Toward an orthogonal central dogma. *Nature Chemical Biology* **2018**, *14*, 103–106

(15) Kassaw, T. K.; Donayre-Torres, A. J.; Antunes, M. S.; Morey, K. J.; Medford, J. I. Engineering synthetic regulatory circuits in plants. *Plant Science* **2018**, *273*, 13–22

(16) Lienert, F.; Lohmueller, J. J.; Garg, A.; Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature reviews Molecular Cell Biology* **2014**, *15*, 95–107

(17) Young, J. W.; Locke, J. C.; Altinok, A.; Rosenfeld, N.; Bacarian, T.; Swain, P. S.; Mjolsness, E.; Elowitz, M. B. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols* **2012**, *7*, 80–88

(18) Belliveau, N. M.; Barnes, S. L.; Ireland, W. T.; Jones, D. L.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Kinney, J. B.; Phillips, R. Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, E4796–E4805

(19) Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell* **1998**, *2*, 437–445

(20) Pichon, X.; Bastide, A.; Safieddine, A.; Chouaib, R.; Samacoits, A.; Basyuk, E.; Peter, M.; Mueller, F.; Bertrand, E. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *Journal of Cell Biology* **2016**, *214*, 769–781

(21) Wu, B.; Eliscovich, C.; Yoon, Y. J.; Singer, R. H. Translation dynamics of single mRNAs in live cells and neurons. *Science* **2016**, *352*, 1430–1435

(22) Wang, C.; Han, B.; Zhou, R.; Zhuang, X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell* **2016**, *165*, 990–1001

(23) Kong, W.; Meldgin, D. R.; Collins, J. J.; Lu, T. Designing microbial consortia with defined social interactions. *Nature Chemical Biology* **2018**, *14*, 821–829

(24) Liao, M. J.; Din, M. O.; Tsimring, L.; Hasty, J. Rock-paper-scissors: Engineered population dynamics increase genetic stability. *Science* **2019**, *365*, 1045–1049

(25) Fiore, G.; Matyjaszkiewicz, A.; Annunziata, F.; Grierson, C.; Savery, N. J.; Marucci, L.; Di Bernardo, M. In-Silico Analysis and Implementation of a Multicellular Feedback Control Strategy in a Synthetic Bacterial Consortium. *ACS Synthetic Biology* **2017**, *6*, 507–517

(26) Jemielita, M.; Wingreen, N. S.; Bassler, B. L. Quorum sensing controls vibrio cholerae multicellular aggregate formation. *eLife* **2018**, *7*, e42057

(27) Khammash, M.; Di Bernardo, M.; Di Bernardo, D. Cybergenetics: Theory and Methods for Genetic Control System. *Proceedings of the IEEE Conference on Decision and Control* **2019**, *2019-December*, 916–926

(28) Sheets, M. B.; Wong, W. W.; Dunlop, M. J. Light-Inducible Recombinases for Bacterial Optogenetics. *ACS Synthetic Biology* **2020**, *9*, 227–235

(29) Rullan, M.; Benzinger, D.; Schmidt, G. W.; Milias-Argeitis, A.; Khammash, M. An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation. *Molecular Cell* **2018**, *70*, 745–756

(30) Chen, S. Y.; Osimiri, L. C.; Chevalier, M.; Bugaj, L. J.; Nguyen, T. H.; Greenstein, R. A.; Ng, A. H.; Stewart-Ornstein, J.; Neves, L. T.; El-Samad, H. Optogenetic Control Reveals Differential Promoter Interpretation of Transcription Factor Nuclear Translocation Dynamics. *Cell Systems* **2020**, *11*, 336–353

(31) Dey, S.; Singh, A. Propagation of stochastic gene expression in the presence of decoys. *Proceedings of the IEEE Conference on Decision and Control* **2020**, *2020-December*, 5873–5878

(32) Razo-Mejia, M.; Marzen, S.; Chure, G.; Taubman, R.; Morrison, M.; Phillips, R. First-principles prediction of the information processing capacity of a simple genetic circuit. *Physical Review E* **2020**, *102*, 022404

(33) Plesa, T.; Stan, G.-B.; Ouldridge, T. E.; Bae, W. Quasi-robust control of biochemical reaction networks via stochastic morphing. *Journal of the Royal Society Interface* **2021**, *18*, 20200985

(34) Munsky, B.; Neuert, G.; Van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **2012**, *336*, 183–187

(35) Neuert, G.; Munsky, B.; Tan, R. Z.; Teytelman, L.; Khammash, M.; Van Oudenaarden, A. Systematic identification of signal-activated stochastic gene regulation. *Science* **2013**, *339*, 584–587

(36) Jones, D. L.; Brewster, R. C.; Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* **2014**, *346*, 1533–1536

(37) Cheong, R.; Rhee, A.; Wang, C. J.; Nemenman, I.; Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **2011**, *334*, 354–358

(38) Kang, T.; Quarton, T.; Nowak, C. M.; Ehrhardt, K.; Singh, A.; Li, Y.; Bleris, L. Robust Filtering and Noise Suppression in Intragenic miRNA-Mediated Host Regulation. *iScience* **2020**, *23*, 101595

(39) Chait, R.; Ruess, J.; Bergmiller, T.; Tkačik, G.; Guet, C. C. Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nature Communications* **2017**, *8*, 1535

(40) Lugagne, J. B.; Sosa Carrillo, S.; Kirch, M.; Köhler, A.; Batt, G.; Hersen, P. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications* **2017**, *8*, 1–7

(41) Guarino, A.; Fiore, D.; Salzano, D.; Di Bernardo, M. Balancing Cell Populations Endowed with a Synthetic Toggle Switch via Adaptive Pulsatile Feedback Control. *ACS Synthetic Biology* **2020**, *9*, 793–803

(42) Fiore, D.; Guarino, A.; Di Bernardo, M. Analysis and control of genetic toggle switches subject to periodic multi-input stimulation. *IEEE Control Systems Letters* **2019**, *3*, 278–283

(43) Fox, Z. R.; Fletcher, S.; Fraisse, A.; Aditya, C.; Sosa-Carrillo, S.; Gilles, S.; Bertaux, F.; Ruess, J.; Batt, G. MicroMator: Open and Flexible Software for Reactive Microscopy. *bioRxiv* **2021**,

(44) Szymańska, P.; Gritti, N.; Keegstra, J. M.; Soltani, M.; Munsky, B. Using noise to control heterogeneity of isogenic populations in homogenous environments. *Physical Biology* **2015**, *12*, 045003

(45) Baumschlager, A.; Aoki, S. K.; Khammash, M. Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. *ACS Synthetic Biology* **2017**, *6*, 2157–2167

(46) Gutenkunst, R. N.; Waterfall, J. J.; Casey, F. P.; Brown, K. S.; Myers, C. R.; Sethna, J. P. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology* **2007**, *3*, 1871–1878

(47) Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **1977**, *81*, 2340–2361

(48) Voliotis, M.; Thomas, P.; Grima, R.; Bowsher, C. G. Stochastic Simulation of Biomolecular Networks in Dynamic Environments. *PLOS Computational Biology* **2016**, *12*, e1004923

(49) Arkin, A.; Ross, J.; McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells. *Genetics* **1998**, *149*, 1633–1648

(50) Bressloff, P. C. Stochastic switching in biology: From genotype to phenotype. *Journal of Physics A: Mathematical and Theoretical 50*, 133001

(51) Freddolino, P. L.; Tavazoie, S. The dawn of virtual cell biology. *Cell* **2012**, *150*, 248–250

(52) Karr, J. R.; Sanghvi, J. C.; MacKlin, D. N.; Gutschow, M. V.; Jacobs, J. M.; Bolival, B.; Assad-Garcia, N.; Glass, J. I.; Covert, M. W. A whole-cell computational model predicts phenotype from genotype. *Cell* **2012**, *150*, 389–401

(53) Kalmar, T.; Lim, C.; Hayward, P.; Muñoz-Descalzo, S.; Nichols, J.; Garcia-Ojalvo, J.; Arias, A. M. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biology* **2009**, *7*, 33–36

(54) Ochiai, H.; Sugawara, T.; Sakuma, T.; Yamamoto, T. Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Scientific Reports* **2014**, *4*, 7125

(55) Gong, S.; Li, Q.; Jeter, C. R.; Fan, Q.; Tang, D. G.; Liu, B. Regulation of NANOG in cancer cells. *Molecular Carcinogenesis* **2015**, *54*, 679–687

(56) Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; Mulligan, V. K.; Dueber, J. E.; Novak, W. R.; El-Samad, H.; Baker, D. De novo design of bioactive protein switches. *Nature* **2019**, *572*, 205–210

(57) Shao, B.; Rammohan, J.; Anderson, D. A.; Alperovich, N.; Ross, D.; Voigt, C. A. Single-cell measurement of plasmid copy number and promoter activity. *Nature Communications* **2021**, *12*, 1475

(58) Gillespie, D. T. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications* **1992**, *188*, 404–425

(59) Mcquarrie, D. A.; Gillespie, D. T. Stochastic Theory and Simulations of Chemical Kinetics. *Journal of Applied Probability* **1967**, *478*, 413–478

(60) Munsky, B.; Khammash, M. The finite state projection algorithm for the solution of the chemical master equation. *Journal of Chemical Physics* **2006**, *124*, 044104