Exploiting noise, nonlinearity, and feedback to differentially control multiple synthetic cells with a single optogenetic input

Michael May[1], Brian Munsky*[1,2],

1 School of Biomedical Engineering, Colorado State University, Fort Collins, CO, USA.
2 Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO, USA.

* brian.munsky@colostate.edu

Abstract

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.

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, or *BioBricks* [2], as simple functional units capable of being mixed and matched to perform desired regulatory needs has yielded powerful capabilities to design cells to perform actions and respond to stimuli [3]. Early advances led to the

development of programmed cells that are capable of advanced responses like switching, complex self-regulation, and fast acting control [4]. Such 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 the design of 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 plant biology and therapeutics [15,16]. Through the use of synthetic biology, plants can be programmed, and certain biological processes of therapeutic interest like protein turnover can be targeted and modified. Improvements to regulatory responses, especially in the form of fluorescent protein reporters [13,17–19], real time signal-gene MCP-MS2-based transcription elongation assays [20] and nascent chain translation elongation assays [13,17–19] has made it possible for cells to transmit their internal states to human observers and respond to RNA and mRNA in addition to protein signals. These technologies allow for more observation and control, not just at the protein level, but at the gene and RNA level as well. Advances in cellular signals have greatly improved the capabilities for different synthetic regulatory modules to communicate with one another and change population dynamics [21,22]. For example, by using a simple model to analyze the effect of cellular quorum sensing on cell densities, simple circuits could be tuned to control cell densities [23].

Although the above advances have been developed primarily to control autonomous biological behaviors, these improvements to regulatory structures, response elements, 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 [24]) paradigm that adds new possibilities for distributed external and internal biological control of synthetic biological circuits [25]. In particular, recent developments in optogenetics [26,27] 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 [28–30]. 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 [31–33], but individual cells appear to behave at random with very little information about their instantaneous external environment [34]. 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 [35], it may seem unlikely that any feedback control strategy could reliably guarantee that a specific cell within a population will respond as desired, and independently of their initial conditions.

Most efforts on external feedback control has focused to manipulate cell population averages [36] or to control individual cells within a larger population [36], both of which can be classified as single-input-single-output (SISO) control. More recent work has sought to control multiple individual cells, each with their own tailored optogenetic inputs [26], which corresponds to multi-input-multi-output control (MIMO). SISO control is limited to control only a single cellular response at a time, while MIMO requires advanced hardware such as digital micromirror devices that devote a separate input to every cell [26]. 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 with 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 sensitivities to small perturbations. 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 all cells see the same light signal.

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 the stochastic model and deterministic SIMO control algorithms to serve as it's 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 target cell. Finally, we show that controls that are designed and optimized using one approximate model of the biological system can be effective to control much more complex biological process whose mechanisms have not been considered during the controller design process.

Materials and Methods

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 $\mathbf{x} = [x_1, \ldots, x_{N'}] \in \mathcal{X}$, 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 state $\mathbf{x}$ is easily concatenated to consider a set of N_c individual cells

$$\mathbf{x} = [x_1^{\text{cell } 1}, \ldots, x_{N'}^{\text{cell } 1}, \ldots, x_1^{\text{cell } N_c}, \ldots, x_{N'}^{\text{cell } N_c}],$$

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 [37]. 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 $u(\mathbf{x}, t)$.

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

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

$$= \mathbf{S}\mathbf{f}(\mathbf{x}, \mathbf{\Lambda}, u(\mathbf{x}, t)), \tag{2}$$

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 (1) can be integrated numerically to describe the system dynamics over time.

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, [38]). Alternatively, one can also use these two properties to uniquely define the Chemical Master Equation (CME, [39]) as:

$$\frac{d}{dt}P(\mathbf{x}) = -\sum_{\mu=1}^{M} w_\mu(\mathbf{x}(t), \mathbf{\Lambda}, u(\mathbf{x}, t))P(\mathbf{x}) + \sum_{\mu=1}^{M} w_\mu(\mathbf{x}-\mathbf{s}_\mu, \mathbf{\Lambda}, u(\mathbf{x}-\mathbf{s}, t))P(\mathbf{x}-\mathbf{s}_\mu). \quad (3)$$

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. 3 is linear in every term $P(\mathbf{x}_i)$, we can now write the CME in matrix format:

$$\frac{d}{dt}\mathbf{P} = \mathbf{A}\mathbf{P}, \quad (4)$$

where the square matrix $\mathbf{A}$ is known as the infinitesimal generator and is defined directly from Eq. 3 as

$$A_{ij} = \begin{cases} -\sum_{\mu=1}^{M} -w_\mu(\mathbf{x}_i, \mathbf{\Lambda}, u(\mathbf{x}_i, t)), & \text{for } i = j, \\ w_\mu(\mathbf{x}_j, \mathbf{\Lambda}, u(\mathbf{x}_j, t)), & \text{for } \mathbf{x}_i = \mathbf{x}_j + \mathbf{s}_\mu \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Because the term $\mathbf{x}_i$ refers to a specific enumerated state vector that is fixed in time, in the special case where $u \equiv u(\mathbf{x}_i)$, the matrix $\mathbf{A}$ is constant with respect to time. For convenience, we can define the control parameter in vector form $\mathbf{u} = [u(\mathbf{x}_1), \ldots, u(\mathbf{x}_N)]$. 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:

$$\frac{d}{dt}\mathbf{P} = (\mathbf{A} + \mathbf{B}(\mathbf{u}))P. \quad (6)$$

In this formulation, although the dynamics of each identical cell was independent and uncoupled in the original infinitesimal generator $\mathbf{A}$, the added infinitesimal generator from the control input, $\mathbf{B}(u)$, introduces coupling between cells. We will see later that this coupling is essential to break symmetry and encourage cells toward desired differential expression phenotypes.

To solve the CME in Eq. 6, we use the Finite State Projection (FSP [40]) 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 x is given by the ordered set $X(x) = [0, 1, ..., n]$ 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 some arbitrary large number. For a system of N_c cells with N' chemically reacting species, where each can reach a maximum of n 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 will be essential, especially if these models are to be used to design optimal control strategies as we discuss below. Further, it will be important to test and verify if control strategies designed and optimized using these simplified models will continue to be effective when applied to more general and more complex systems.

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 [41] to constrain each of these models. 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 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 their 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 the more complicated system with different and hidden dynamics.

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

To introduce single-cell stochastic feedback control, we develop an unregulated model (denoted as $\mathcal{M}_U$) to describe the light-induced activation of an optogenetically controlled T7 polymerase. As depicted in Fig.1(A), this system contains two light activated T7 domains ($T7_n$ and $T7_c$) that are produced at constant rates, and which degrade in a first order decay process. These domains dimerize when subject to light activation leading them to form an active T7 polymerase. The T7 polymerase dimer can bind to, and activate, a gene to initiate protein production (transcription and translation are lumped into a single event). Proteins are assumed to degrade according a first order rate process. These interactions are described by the following reactions

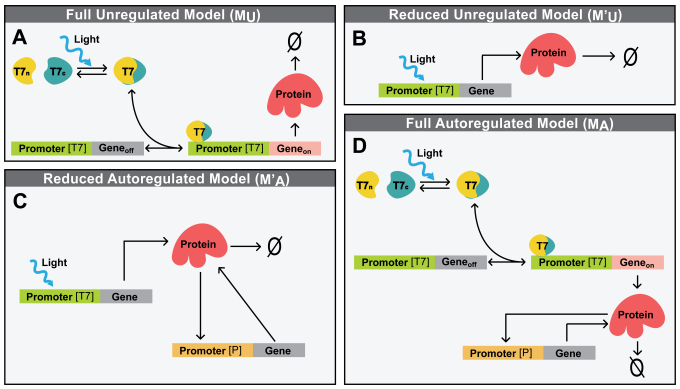

Fig 1. Diagrams of models for light-induced gene regulation with and without auto-regulation. (A) The full physical model ($\mathcal{M}_U$) describes the full mechanistic processes of T7 polymerase dimerization and gene activation in addition to protein production and degradation. (B) A simplified model ($\mathcal{M}'_U$) 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 a secondary auto-regulation promoter.

$$\mathcal{M}_u = \left\{ \begin{array}{rcl} \emptyset & \underset{k_n}{\overset{\gamma_n[T7_n]}{\rightleftharpoons}} & T7_n, \\ \emptyset & \underset{k_c}{\overset{\gamma_c[T7_c]}{\rightleftharpoons}} & T7_c, \\ T7_n + T7_c & \underset{u(\phi)T7_nT7_c}{\overset{k_{i1}T7}{\rightleftharpoons}} & T7\ , \\ T7 + g_{\text{off}} & \underset{k_{f2}T7g_{off}}{\overset{k_{i2}g_{on}}{\rightleftharpoons}} & g_{\text{on}}\ , \\ g_{\text{on}} & \xrightarrow{k_{f1}[g_{on}]} & P + g_{\text{on}}\ , \\ P & \xrightarrow{\gamma} & \emptyset\ , \end{array} \right\} \tag{7}$$

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

$$\frac{d}{dt}\begin{bmatrix} T7_n \\ T7_c \\ T7 \\ g \\ P \end{bmatrix} = \begin{bmatrix} -u(\phi)[T7_n][T7_c] + k_{i1}[T7] + k_n - \gamma_n T7_n \\ -u(\phi)[T_7n][T7_c] + k_{i1}[T7] + k_c - \gamma_n T7_n \\ u(\phi)[T7_n][T7_c] - k_{i1}[T7] - k_{f2}[T7](g_{total} - [g]) + k_{i2}[g] \\ k_{f2}[T7](g_{total} - [g]) - k_{i2}[g] \\ k_{f3}[g] - \gamma[mRNA] \end{bmatrix}. \tag{8}$$

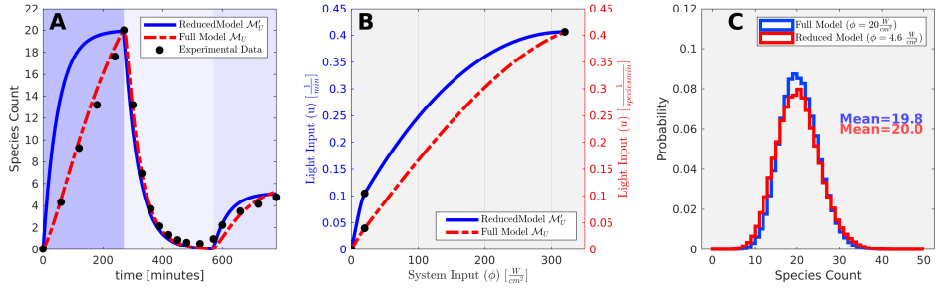

Fig 2. Parameter estimation of unregulated models $\mathcal{M}'_U$ and $\mathcal{M}_U$. (A) Fits of $\mathcal{M}_U$ (red line) and $\mathcal{M}'_U$ (blue line) to experimental data from Baumschlager et al. [41] (black dots).(B) Calibration curves representing the conversion of light [Watts per cm^2] 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.

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 [41], in which the system was subjected to three different levels of UV radiation and allowed to respond over time (see Materials and Methods for fitting procedures). The resulting spline fit to the data is shown by the red dashed lines in Fig. 2A. In the model, the relationship between light input (units of Watts per cm squared) and the T7 dimerization rate (units of species per second) is assumed to have a general spline fit form, which is depicted by the red lines in Fig. 2B.

Having determined a baseline ODE-based model that yields a good fit to existing experimental temporal data for the system, 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 from [41], 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 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 $u(\phi)$ is a nonlinear function of the light intensity (ϕ). Under this simple rule, the

approximate dynamics of $P(t)$ are written simply as:

$$\frac{dP}{dt} = u(\phi) - \gamma P + \eta \tag{9}$$

Under this approximation model, which we denote as $\mathcal{M}'_U$, the model parameter, γ and the $u(\phi)$ calibration curves are simultaneously fit to capture the same time dynamics of the experimental data from [41]. The calibration curve was parameterized using a spline function and fit to match the experimental data points with the expected system input. The resulting fit of model $\mathcal{M}'_U$ to the data is shown in the solid blue line Fig. 2A, and the identified calibration curve for $u(\phi)$ is shown by the solid blue line in Fig. 2B. All parameters for the reduced model are shown in Table .

Parameter Name	Parameter Value
γ	0.0203
η	0.0001

We next extended both the original model and the simplified model to include discrete stochastic events for protein production and degradation, as well as for T7 dynamics for the full model. Fig. 2C shows the results of twenty four stochastic simulations at a long time for calibrated inputs that result in a mean of twenty protein molecules per cell using the full model (red) or the simplified model (blue). Although both models provide reasonably good, but not identical, matches the mean protein production level and the overall time scales as shown in Fig. 2A, we observed similar behavior in their distributions despite the differences in model architecture. It remains to be seen if the differences in architecture and time scales need to be accounted for when we use the simplified model 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.

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 the possibility of 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. 1.C shows a schematic of the new model (which we denote as $\mathcal{M}'_A$) that has auto-regulation due to the addition of a secondary promoter that 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:

$$\frac{dP}{dt} = u(\phi) + \eta + k_o \frac{x^\mu}{x^\mu + \beta^\mu} - \gamma P, \tag{10}$$

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

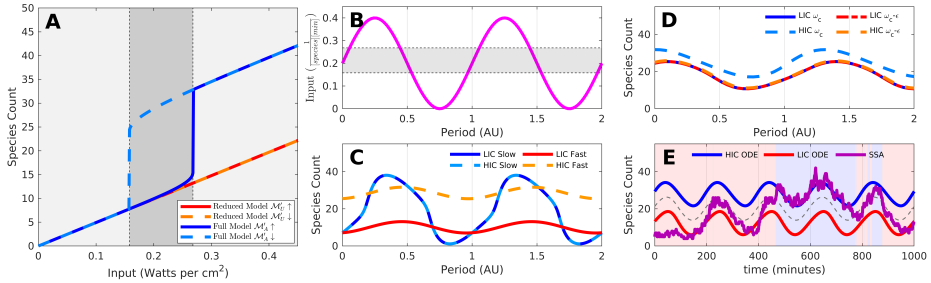

Fig 3. Input-output analysis for the simplified model with and without auto-regulation (Models $\mathcal{M}'_U$ and $\mathcal{M}'_A$). (A) Bifurcation (hysteresis) analyses of $\mathcal{M}'_U$ (red shades) and $\mathcal{M}'_A$ (blue shades). Dashed (solid) lines depict the change in the steady state 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 crosses and circles) only occur for the $\mathcal{M}'_A$. (B) Input used for sinusoidal input excitation. (C) Steady state temporal behavior of $\mathcal{M}'_A$ under sinusoidal input. 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 RMP, red shades), slow (0.001 RMP, blue shades). All plots show steady state temporal response after two oscillation periods or 3500 minutes, whichever is longer. (D) Same as (C) but with frequencies 0.0043522 RMP (red shades) and 0.0043525 RMP (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 analyses using Gillespie's SSA [38].

auto-regulation promoter with leakiness η, and the final term corresponds to the first order decay of the protein. The new model parameters Hill parameters (β, μ) 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 .

Parameter Name	Parameter Value
κ	0.406
γ	0.0203
β	20
μ	8
η	0.0001

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 becomes apparent in the auto-regulated model, but is not observed in the unregulated model. Low and high light 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 remain in 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 flight 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 will be more interested in how cells might respond to light fluctuations at faster, transient time scales. Therefore, we next test the stability of input-to-output behaviors under time-varying input signals. We start all simulations with two cells with identical parameters, but at different initial conditions (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 that sweeps a input range that 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 all transient dynamics have had time to decay, and the time axis is scaled to show the response over two oscillation periods. Fig. 3C shows in red, 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 (blue 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 $\omega_c - \varepsilon$ but is lost at a slightly slower frequency of $\omega_c + \varepsilon$, where $\omega_c = 0.0043525$ RMP is the critical frequency and $\varepsilon = 3e - 7$ RMP is a small perturbation to that frequency.

Using bistable deterministic analyses, 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 know 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. 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 cell to or from separate phenotypes. Specifically, we compared the ODE and the SSA analysis of $\mathcal{M}'_U$ under a sinusoidal input with moderate input frequency of $\omega = 3e^{-7} > \omega_c \; min^{-1}$ 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. In the next section we will explore how this condition-dependent disruption of symmetry could in principle be exploited at favorable points in time to push one cell to a chosen phenotype while forcing another cell or group of cells toward a different cell fate. In the following section, we explore control strategies to exploit this effect.

Finite State Projection analysis and optimization can be combined to design optimal control strategies that control two cells using a single global time-varying input.

As described in Methods and Materials, the master equation for a finite number of cells subject to a fluctuating environmental input signal can be written as

$$\frac{d}{dt}\mathbf{P} = (\mathbf{A} + \mathbf{B}(\mathbf{u}(t)))\mathbf{P}. \tag{11}$$

where $\mathbf{A} \in \mathbb{R}^{N_c N' \times N_c N'}$ is a block diagonal infinitesimal generator that has one $\mathbb{R}^{N' \times N'}$ block for each of N_c independent cells; $\mathbf{u} \in \mathbb{R} \to \mathbb{R}^{N_c N' \times 1}$ is a (potentially time varying) vector of control inputs with one entry for every possible state in state space; $\mathbf{P} \in \mathbb{R}_{\geq 0}^{N_c N' \times 1}$ is the probability mass vector for all possible states; and $\mathbf{B}$ is an $(\mathbb{R}^{N_c N' \times 1} \to \mathbb{R}^{N_c N' \times N_c N'})$ tensor operation that accepts control inputs and creates a coupled infinitesimal generator matrix to define the stochastic action of the controller. At first, we will consider the special case where $\mathbf{u}$ depends only on the definition of the enumerated states; in this case, both $\mathbf{u}$ and $\mathbf{B}(\mathbf{u})$ are constant with respect to time and Eq. (11) is the master equation for a Markov process.

We consider two cells, for which the fluctuating vector $\tilde{\mathbf{X}} = [\tilde{x}_1, \tilde{x}_2]^T$ denotes the instantaneous state of the protein levels for these cells. We then specify an arbitrary target state that we wish to control these cells to reach (e.g., $\hat{\mathbf{T}} = [\hat{T}_1, \hat{T}_2]^T = [30, 10]^T$), corresponding to the first cell in the high expression phenotype and the second cell in the low expression phenotype, and we define a cost function as the expected squared distance, $J = \mathbb{E}\{||\tilde{\mathbf{X}} - \hat{\mathbf{T}}||_2^2)$. This expected value can be written as

$$\begin{aligned} J &= \mathbb{E}\{||\mathbf{X} - \mathbf{T}||_2^2) && (12)\\ &= \sum_{i,j} P(\tilde{x}_1 = i, \tilde{x}_2 = j)((i - \hat{T}_1)^2 + (j - \hat{T}_2)^2) && (13)\\ &= \sum_{i,j} P_{ij} C_{ij} = \mathbf{CP}, && (14) \end{aligned}$$

where $\mathbf{C}$ is the vector of squared distances of each state from the target. With this formal definition of the cost function J and the CME from Eq. 11, we can then compute its gradient with respect to the control law as:

$$\frac{\partial J}{\partial u} = \mathbf{C}(\mathbf{A} + \mathbf{Bu})^{-1}(\mathbf{BMP}), \tag{15}$$

(see derivation in SI). In this formulation of the cost function gradient with respect to the controller, we have introduced a matrix $\mathbf{M}$ to stand in for different potential control assumptions. By changing the specification of $\mathbf{M}$, we can consider many different control rules, but here we consider only three different possibilities: an unaware controller (UAC) that has no information about either cell to make its control decision; a fully aware controller (FAC) which has complete knowledge of x and y and therefore uses the full space of $u(x, y)$ to make its control decisions; and a partially-aware controller (PAC) which uses only information of a single cell, y, to make its control decisions $u(y)$. See SI for explicit formulation of $\mathbf{M}$ for each of these controllers.

Having specified the cost function gradient for each controller, we can run an optimization algorithm to search along the gradient to find local minima for the cost function. We reiterate 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 the optimal control strategy in the form of a scalar field $u(x, y)$, but 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 $u(x, y)$ as a function of the species quantities x for the first cell and y 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 $u(x, y)$ on the system. 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' distribution is made tighter at some intermediate value between the two target values for cell one and cell two; this results in a slightly better numerical cost value, but it becomes even less both cell 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, 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 control of multiples cells using a single signal. Specifically, noise breaks the symmetry of cell behavior and allows cells to switch independently between phenotypes, feedback

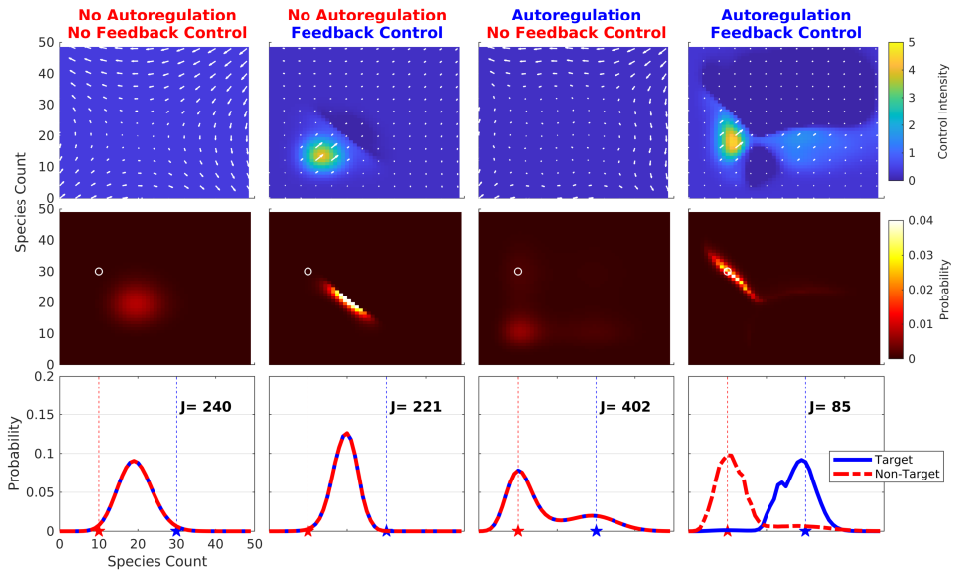

Fig 4. Optimized control laws and performance results for cells and different combinations of regulatory structure and control strategy. (Top row) Optimized control input versus cells' states, $\mathbf{u}(x, y)$ (colormap at far right). White vectors plotted over the control inputs represent the net flow of probability at that point in state space. (Middle row) Resulting joint probability distribution ($\mathbf{P}(x, y)$, colormap at far right). The target point $T = [30, 10]^T$ is denoted by a white circle. (Bottom row) Corresponding marginal distributions for target cell ($\mathbf{P}(y)$, blue) and non-target cell ($\mathbf{P}(x)$, 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 columns shows the results for auto-regulation and fully aware feedback control (FAC in text).

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 target cell to one state and all non-target cells to another state. To handle this problem, we first utilize a partially aware controller (PAC) that observes only the first target cell and ignores the states of all other cells. Fig. 5 compares the results of this simplified controller (middle column) to that of the two-cell controller from the previous section (left column). The resulting control law, $u(y)$, is optimized to find the best control signal input for each possible value y of the target cell, and the top row of Fig. 5, we see that this optimal PAC controller depends only on the observed target cell (y axis), but is constant with respect

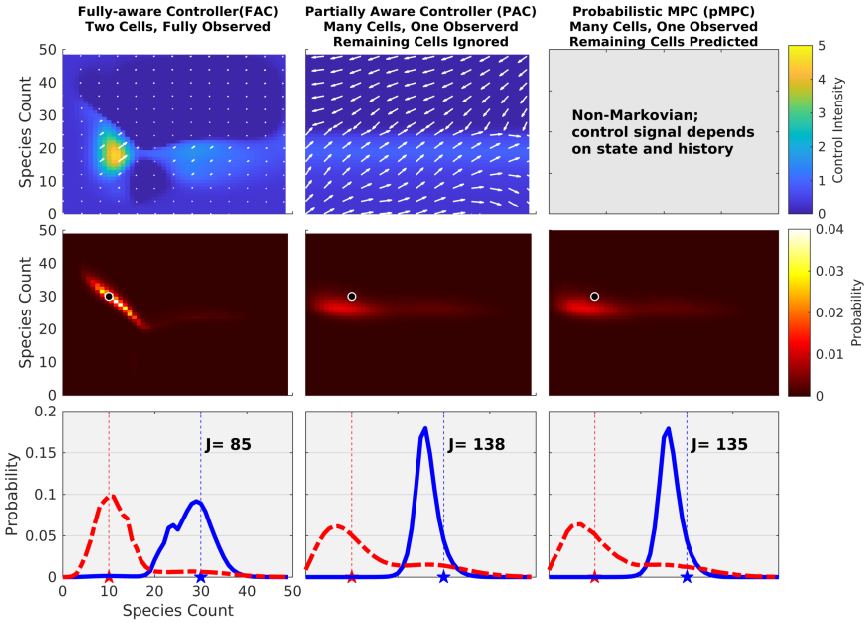

Fig 5. Control laws and performance for strategies to control many cells at once using partial knowledge only of a single target cell. Top, middle and bottom rows show the control law $u(x, y)$, the resulting joint probability distribution ($\mathbf{P}(x, y)$), and the marginal distributions ($P(y)$, blue) and non-target cell ($\mathbf{P}(x)$, red). The middle column shows results for the partially aware controller (PAC in text), where the control depends on a single target cell ($\mathbf{u}(y)$) and any arbiltrarilly large number of non-target cells (x). 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 ($\mathbf{u}(y, \mathbf{P}(x))$ depends not only upon the state of observed y, but also on the predicted non-target probability vector $\mathbf{P}(x)$. 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.

to all non-target cells (x axis). Despite the simplicity of this control strategy and the fact that it requires only knowledge of the instantaneous expression of the target cell, the second and third rows of Fig. 5 shows that the PAC controller effectively breaks symmetry to force the target cell to a high expression phenotype while most non-target cells are correctly directed to the low expression phenotypes. Although the partially aware controller slightly under performs 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 non-target cells (Supplemental Fig. S1). In contrast, to use the FAC for more than one non-target cell requires modification (e.g., training of a higher rank tensor representation of the control algorithm, or defining to control based on the mean, median, or some other statistical quantity for the non-target cells), and rapidly loses performance as the number of cells is increased. Indeed, for any more than a single non-target cells, the PAC outperforms the FAC (Supplemental Fig. S1).

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 approach controlled based on only the observation of a single target cell, and had no information about the other non-target cells. However, knowing the history of the input signal (i.e., the light intensity over time from the past), the FSP approach allows for the possibility that the controller can predict probability distribution of all non-target cells. With this possibility in mind, we next explored a new class of controller that could use direct knowledge of the protein expression in the target cell, the known control input signal at the current time, and a probabilistic model to predict the probability distribution of expression in the unmeasured cells. Specifically, we use the FSP approach to integrate our prediction of the probability distribution for the non-target cells as:

$$\frac{d}{dt}\mathbf{P}_{nt} = \left(\mathbf{A} + \hat{\mathbf{B}}\tilde{u}(t)\right)\mathbf{P}_{nt}, \tag{16}$$

where $\mathbf{P}_{nt} \in \mathbb{R}_{\geq 0}^{N' \times 1}$ is the probability mass vector for the protein expression in the non-target cells, $\mathbf{A} \in \mathbb{R}^{N' \times N'}$ is the infinitesimal generator for a single cell in the absence of any control input, $\hat{\mathbf{B}} \in \mathbb{R}^{N' \times N'}$ is a banded matrix with negative one on the diagonal ($\hat{B}_{i,i} = -1$) and positive one on plus-one diagonal ($\hat{B}_{i+1,i} = 1$), and $\tilde{u}(t)$ is the instantaneous input signal that is produced by the controller.

Using this prediction for the non-target cell, the pMPC controller law can now be defined as

$$u(y,t) = u(y) + Z\mathbf{P}_{nt}(t) \tag{17}$$

where $u(y)$ is the PAC optimized controller, and Z is a set of linear weights which adjusts the input based off of the estimated probability distribution $\mathbf{P}_{nt}$. Rather than seek to optimize this controller, we simply take the existing PAC control strategy from before, $u(y)$ and add a correcting term which adjusts the PAC to account for the predicted distribution information and show it outperforms the PAC alone.

Fig. 5 shows that the resulting *non-optimized* pMPC controller outperforms the *fully optimized* PAC. Further optimization of the pMPC approach to (i) choose the right statistical quantity or quantities to use for s_x, or (i) improve the control law $u(s_x, y)$, would certainly lead to improvements in the control performance. However, 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.

For a closer look at how the pMPC approach works to control target and non-target cells alike, Figs. 6A shows an example control input over time and Fig. 6B shows the resulting trajectories over time for a target cell (blue), the predicted probability distribution for non-target cells (gray shading), and a single unobserved non-target cell (red). We reiterate that the controller has no direct knowledge of the red line. From the figure, it is clear that target cell is well controlled with low variability. Moreover, Fig.

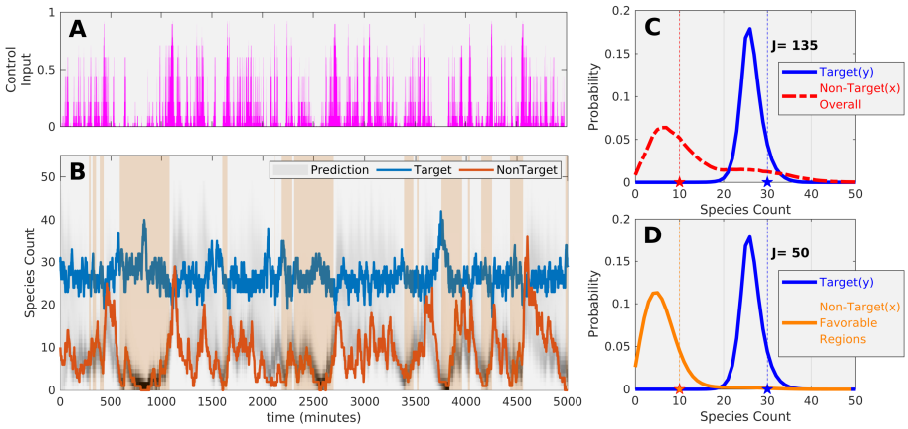

Fig 6. Control signal and response versus time for pMPC control. (A) Control signal generated by the pMPC controller. (B) Predicted transient distributions for non-target cells (gray), target cell (blue), and a single non-observed non-target cell (red). (C) Time averaged performance of the control law in terms of marginal distributions for the target (blue) and non-target (red) cells. (D) Average of the pMPC performance, when considering only periods in time when the model predicts success in the sense that the non-target cell is has low expression ($x < 15$).

6B shows that knowledge of the fluctuating input signal is sufficient to yield good predictions of the non-target cell response (compare red line with dark gray shading), although as expected there are periods of poor predictions when the specific non-target cell samples the higher or lower tail of the predicted distribution. In addition to outperforming the PAC approach in terms of the overall cost function, the pMPC provides additional knowledge predictions for when the controller is effective, and when non-target cells are more likely to escape from their intended phenotype. To illustrate this Figs. 6C shows the marginal distributions averaged over all times, and Fig. 6D shows the marginal distribution only for the periods of time when the probabilistic model predicts that 90% or more of the non-target cells have less than 15 molecules per cell (these regions are denoted in orange in Fig. 6B). By focusing only on these times identified as successful by the controller, the cost function of the controller substantially decreases from $J = 135$ to $J = 50$. These results suggest that predicted dynamic information about the non-target cells can not only be used to improve the quality of the controller, but that it the pMPC can also be used to self-identify when it is working well and when it is not.

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

We next ask how well would the controllers designed above using simplified stochastic dynamics models work if they were applied to control more complex systems that contain additional hidden states and which have unknown dynamics and time delays. For this analysis, we first used steady state ODE analyses of the full auto-regulation

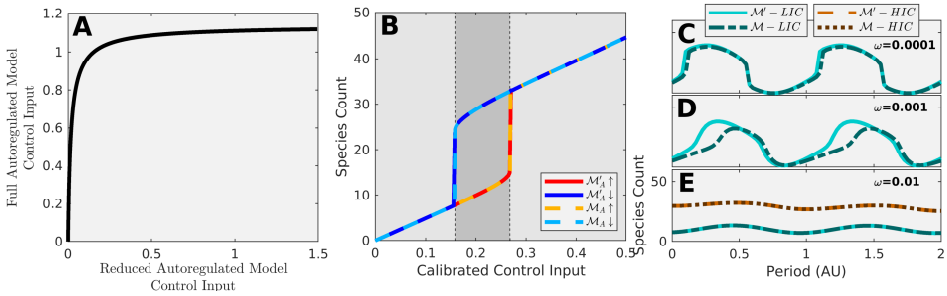

Fig 7. Calibration and use of controllers for use with a new, more complex model. (A) Calibration curve identified to match steady state ODE of simple and complex auto-regulation model ($\mathcal{M}'_A$ and $\mathcal{M}'_A$ from Fig. 1C,D). (B) Steady state analyses shows that the calibrated inputs result in similar hysteresis behavior for both $\mathcal{M}'_A$ and $\mathcal{M}_A$. (C) Input-output response analysis showing that $\mathcal{M}_A$ with calibrated inputs closely matches behavior of $\mathcal{M}'_A$ at slow frequencies (0.0001 RMP), but (D) the more complex model begins to lag at moderate frequencies (0.001 RMP). (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.

model (Fig. 1D) and the reduce auto-regulation models (Fig. 1C) to define an input calibration curve between the two models, the results of which are shown in Fig. 7A. After calibration, we verified that the full and reduced auto-regulation models resulted in similar bifurcation diagrams as shown in Fig. 7B. However, although calibration allowed us to match the steady state and slowly fluctuating input responses of the two models, the noise and temporal responses to fast inputs are qualitative and quantitatively different, as can be observed by very different distributions in Figs. 2C and different input-to output time lags or amplitudes in Figs. 7D,E.

Having calibrated the controller for the full model to match the response of the simpler model, we took the UAC, FAC, and PAC controllers from above applied them directly (i.e., without any further tuning or optimization) to the full mechanistic model with auto-regulation. However, despite the differences in 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

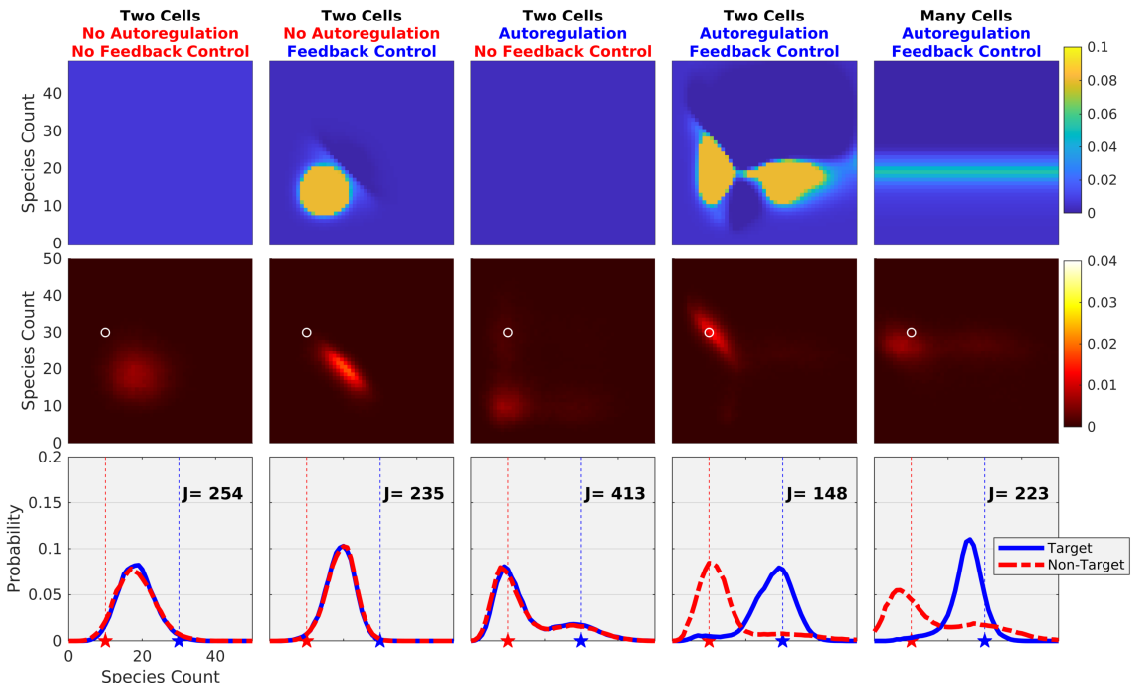

Fig 8. Full Models are paired with the calibrated controllers (top row) to solve for the joint probability distributions (middle row) and marginal probability distributions (bottom row). As before both auto-regulation and feedback are needed to break symmetry and control fails if either of these are 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 target cell and an arbitrary number of non-target cells.

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.

We have also demonstrated three new types of optimizable stochastic controllers that rely on the integration of noise, non-linear auto-regulation, and chemical or optogenetic feedback, and which allow for the simultaneous control of multiple cells using a single external input signal. The first of these three, 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 target 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 [42], 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 with qualitatively

and quantitatively different dynamic and stochastic 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, 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 unlikely to differentiate [43]. As such, it might be interesting to explore the possibility that temporally controlled fluctuations in Nanog transcription factors **??** 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 [12,44], improved live cell reporters [13], and faster and more specific optogenetically controlled transcription factors inputs [27], 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.

Supporting information

1 Supplemental

This section contains information about optimization of model-controller pairs, and some additional figures and analyses.

1.1 Model Predictive Control

The tensor form of the FSP can be given by:

$$\frac{\partial P^i}{\partial t} = (S^i_{jm} r^m + B^i_{jm} u^m) P^j \tag{18}$$

Where S and B are the stoichiometry connection tensor of the system and the controller, $\mathbf{r}^m$ the vector of system reactions and $\mathbf{u}^m$ the vector of control input reactions. The generation of any stoichiometry connection is given by

$$\mathbf{S}^i_{jm} = \begin{cases} 1, & \text{for } \mathbf{x}_i = \mathbf{x}_j + \mathbf{s}_m \text{ over all m reactions} \\ -1, & \text{for } \mathbf{x}_i = \mathbf{x}_j - \mathbf{s}_m \text{ over all m reactions} \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

Where S is the connection tensor and $\mathbf{s}_m$ is the stoichiometry vector of the m'th reaction. The S tensor and B tensor form the connections between nodes of a state space due to a particular controllable reaction u^m over all m and changes only with the stoichiometry of a reaction set. The product given by the contraction of $S^i_{jm} r^m$ with r^m generates the infinitesimal generator created by all reactions of $\mathbf{r}$. Likewise the contraction of the B^i_{jm} tensor with $\mathbf{u}^m$ generates the infinitesimal generator given by the control reactions $\mathbf{u}$. Deriving how P changes with small changes in $\mathbf{u}$ gives

$$\frac{\partial P^j}{\partial u^n} = P^j_{,n} = inv(Sr + Bu)^j_i B^i_{ko} M^o_n P^k \tag{20}$$

Where M changes depending on the controller type. If M is the identiy matrix and $\mathbf{u}$ is the the full size, the then optimization works for the FAC. If M is an m by 1 vector of ones and $\mathbf{u}$ is a vector with a single element, then it works for the UAC. For M as rectangular matrix of appropriately placed ones and zeros and $\mathbf{u}$ of size $\mathbf{R}^{N_c}$, it optimizes the PAC. Finally, plugging this into the score we have:

$$\frac{\partial J}{\partial u^n} = \frac{\partial (C_j P^j)}{\partial u^n} = C_j P^j_{,n} = C_j inv(Sr + Bu)^j_i B^i_{ko} M^o_n P^k \tag{21}$$

Then $\mathbf{u}$ can be optimized by starting at an appropriate $\mathbf{P}$ and changing $\mathbf{u}$ in the direction of $J_{,n}$, and then updating the new $\mathbf{P}$.

1.2 Model Predictive Control

Model predictive control uses a simple linear machine to generate light inputs for $\mathbf{u}$ such that

$$u(y,t) = c(y) + Z\mathbf{P}_{nt}(t) \tag{22}$$

Where Z is a simple linear machine which takes the probability of the non-target probability distribution $\mathbf{P}_{nt}$ and outputs the light output. The set of weights used to describe A are shown in (Fig. 9A).

1.3 Supplemental Analysis

The analysis performed in this paper sometimes yielded scores with high variability between simulations and large changes in score despite remarkable similarities in the distributions of the solutions. To analyze these characteristics, SSA simulations were performed using a set of controllers and the score at each time point measured. Histograms of score over time for an entire time-series trajectory using the FAC, PAC, and pMPC show a long tail distribution of the score (Fig. 9B), suggesting that these behaviors observed are due to long tailed behavior of the score. These data taken

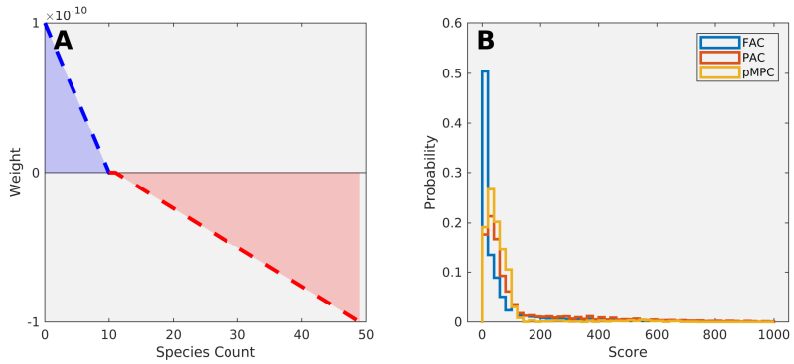

Fig 9. (A) Weights of Z show that the pMPC tends to increase the controller when $\mathbf{P}_n t$ is weighted below ten, but tends to decrease the control signal when $\mathbf{P}_n t$ is weighted above ten. (B) Distribution of scores obtained during time trajectory show that score over time is a heavy tailed distribution. Although the probability of a high score is low, the score value itself tends to be vary large and can increase variability in simulations as well as attributing large differences in score for similar looking distributions.

together suggest that the score is often low, but has rare moments in time where the performance is exceedingly poor, causing a large temporary increase in score and raising the average despite overall good performance.

A set of five stochastic controllers are simulated thirty two times each using two, four, eight, and sixteen non-target cells to show the effect a varying number of non-target cells has on the control performance of the systems. The five controllers chosen are to be the FAC, PAC, and pMPC in addition to the mean input of the FAC controller over all non-targets (MFAC) and the FAC light input on the mean expression of all non-target cells (FACM). The data shown in Fig.10.A shows the median expression in the colored line as well as the 25th and 75th percentile in the colored region. These data in Fig.10 show that the performance of the FAC , MFAC, and FACM decrease as the number of non-target cells increases, but the PAC and pMPC performance remains steady as the non-target cells increase. Although the FACM has similar median performance, its variability is larger than the FAC and MFAC. Although performance decreases in the FAC, MFAC and FACM, they still perform much better than the UAC. These data taken together suggest that the PAC and pMPC are both better controllers once the number of non-target cells is greater than or equal to two, and attempts to extend the FAC to work on many cells is not as good as optimizing unique controllers for each. The score of the FAC and PAC are similar at two non-target cells (144 and 141 respectively) (Fig. 10A) and (Fig. 10D), but at 16 non-targets the score of the FAC degrades (138), while the PAC remains the same(137) (Fig. 10C) and (Fig. 10E). Additionally, shape changes can be observed in the FAC from two to sixteen non-target cells, while the PAC remains the exactly same.

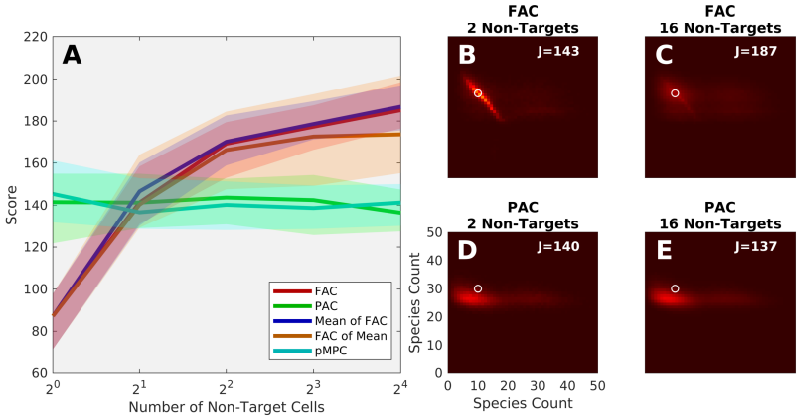

Fig 10. Stochastic controllers simulated using a varying number of non-target cells show some controllers are dependent on the number of non-targets while others are not. (A) median score of 32 simulations using a set of five controllers, with 25% and 75% quartiles shown in the colored banded region. The FAC, MFAC, and FACM controllers all show dependence on the number of controllers, while the PAC and pMPC show no such dependence. The data presented shows FAC is superior with a single non-target cells, while the PAC and pMPC are superior with many non-target cells. (B and C) FAC controller joint distribution of the target and the non-targets for two cells and sixteen cells respectively. (D and E) PAC controller joint distribution of the target and the non-targets for two cells and sixteen cells respectively shows no change in the joint distribution.

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.

References

1. Hasty J, McMillen D, Collins JJ. Engineered gene circuits. Nature. 2002;420:224 EP –.

2. Knight T. Idempotent Vector Design for Standard Assembly of Biobricks. MIT Libraries. 2003; p. 1–11.

3. Jang SH, Cha JW, Han NS, Jeong KJ. Development of bicistronic expression system for the enhanced and reliable production of recombinant proteins in Leuconostoc citreum. Scientific Reports. 2018;8(1):1–11. doi:10.1038/s41598-018-27091-z.

4. Menolascina F, Fiore G, Orabona E, De Stefano L, Ferry M, Hasty J, et al. In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks. PLOS Computational Biology. 2014;10(5):1–14. doi:10.1371/journal.pcbi.1003625.

5. Gardner T, Cantor C, Collins J. Construction of a Genetic Toggle Switch in Escherichia coli. Nature. 2000;403:339–342. doi:10.1038/35002131.

6. Wu M, Su RQ, Li X, Ellis T, Lai YC, Wang X. Engineering of regulated stochastic cell fate determination. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(26):10610–10615. doi:10.1073/pnas.1305423110.

7. Kramer BP, Viretta AU, Daoud-El Baba M, Aubel D, Weber W, Fussenegger M. An engineered epigenetic transgene switch in mammalian cells. Nature biotechnology. 2004;22(7):867–870.

8. Schaumberg KA, Antunes MS, Kassaw TK, Xu W, Zalewski CS, Medford JI, et al. Quantitative characterization of genetic parts and circuits for plant synthetic biology. Nature Methods. 2016;13(1):94–100. doi:10.1038/nmeth.3659.

9. Shin J, Zhang S, Der BS, Nielsen AA, Voigt CA. Programming Escherichia coli to function as a digital display . Molecular Systems Biology. 2020;16(3):1–12. doi:10.15252/msb.20199401.

10. Ng AH, Nguyen TH, Gómez-Schiavon M, Dods G, Langan RA, Boyken SE, et al.. Modular and tunable biological feedback control using a de novo protein switch; 2019.

11. Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA. Genetic programs constructed from layered logic gates in single cells. Nature. 2012;491:249 EP –.

12. Groseclose TM, Rondon RE, Herde ZD, Aldrete CA, Wilson CJ. Engineered systems of inducible anti-repressors for the next generation of biological programming. Nature Communications. 2020;11(1):1–15. doi:10.1038/s41467-020-18302-1.

13. Morisaki T, Lyon K, DeLuca KF, DeLuca JG, English BP, Zhang Z, et al. Real-time quantification of single RNA translation dynamics in living cells. Science. 2016;352(6292):1425–1429. doi:10.1126/science.aaf0899.

14. Liu CC, Jewett MC, Chin JW, Voigt CA. Toward an orthogonal central dogma. Nature Chemical Biology. 2018;14(2):103–106. doi:10.1038/nchembio.2554.

15. Kassaw TK, Donayre-Torres AJ, Antunes MS, Morey KJ, Medford JI. Engineering synthetic regulatory circuits in plants. Plant Science. 2018;273(April):13–22. doi:10.1016/j.plantsci.2018.04.005.

16. Lienert F, Lohmueller JJ, Garg A, Silver PA. Synthetic biology in mammalian cells: next generation research tools and therapeutics. Nature reviews Molecular cell biology. 2014;15(2):95–107.

17. Pichon X, Bastide A, Safieddine A, Chouaib R, Samacoits A, Basyuk E, et al. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. Journal of Cell Biology. 2016;214(6):769–781. doi:10.1083/jcb.201605024.

18. Wu B, Eliscovich C, Yoon YJ, Singer RH. Translation dynamics of single mRNAs in live cells and neurons. Science. 2016;352(6292):1430–1435. doi:10.1126/science.aaf1084.

19. Wang C, Han B, Zhou R, Zhuang X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell. 2016;165(4):990–1001. doi:10.1016/j.cell.2016.04.040.

20. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. Molecular cell. 1998;2(4):437–445.

21. Kong W, Meldgin DR, Collins JJ, Lu T. Designing microbial consortia with defined social interactions. Nature Chemical Biology. 2018;14(8):821–829. doi:10.1038/s41589-018-0091-7.

22. Liao MJ, Din MO, Tsimring L, Hasty J. Rock-paper-scissors: Engineered population dynamics increase genetic stability Michaelcrease Genetic Stability. Science. 2019;365(September):1045–1049.

23. Jemielita M, Wingreen NS, Bassler BL. Quorum sensing controls vibrio cholerae multicellular aggregate formation. eLife. 2018;7:1–25. doi:10.7554/eLife.42057.

24. 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(Cdc):916–926. doi:10.1109/CDC40024.2019.9030209.

25. Sheets MB, Wong WW, Dunlop MJ. Light-Inducible Recombinases for Bacterial Optogenetics. ACS Synthetic Biology. 2020;9(2):227–235. doi:10.1021/acssynbio.9b00395.

26. Rullan M, Benzinger D, Schmidt GW, Milias-Argeitis A, Khammash M. An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation. Molecular Cell. 2018;70(4):745–756.e6. doi:10.1016/j.molcel.2018.04.012.

27. Chen SY, Osimiri LC, Chevalier M, Bugaj LJ, Nguyen TH, Greenstein RA, et al. Optogenetic Control Reveals Differential Promoter Interpretation of

Transcription Factor Nuclear Translocation Dynamics. Cell Systems. 2020;11(4):336–353.e24. doi:10.1016/j.cels.2020.08.009.

28. 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-Decem(Cdc):5873–5878. doi:10.1109/CDC42340.2020.9303901.

29. 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(2):1–14. doi:10.1103/PhysRevE.102.022404.

30. Plesa T, Stan GB, Ouldridge TE, Bae W. Robust control of biochemical reaction networks via stochastic morphing. arXiv. 2019;.

31. Munsky B, Neuert G, Van Oudenaarden A. Using gene expression noise to understand gene regulation. Science. 2012;336(6078):183–187. doi:10.1126/science.1216379.

32. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, Van Oudenaarden A. Systematic identification of signal-activated stochastic gene regulation. Science. 2013;339(6119):584–587. doi:10.1126/science.1231456.

33. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014;346(6216):1533–1536. doi:10.1126/science.1255301.

34. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. Information transduction capacity of noisy biochemical signaling networks. science. 2011;334(6054):354–358.

35. Kang T, Quarton T, Nowak CM, Ehrhardt K, Singh A, Li Y, et al. Robust Filtering and Noise Suppression in Intragenic miRNA-Mediated Host Regulation. iScience. 2020;23(10). doi:10.1016/j.isci.2020.101595.

36. Chait R, Ruess J, Bergmiller T, Tkačik G, Guet CC. Shaping bacterial population behavior through computer-interfaced control of individual cells. Nature communications. 2017;8(1):1–11.

37. Gillespie DT. A rigorous derivation of the chemical master equation. Physica A: Statistical Mechanics and its Applications. 1992;188(1-3):404–425. doi:10.1016/0378-4371(92)90283-V.

38. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry. 1977;81(25):2340–2361.

39. Mcquarrie DA, Gillespie DT. Stochastic Theory and Simulations of Chemical Kinetics. Journal of Applied Probability. 1967;478(1967):413–478.

40. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. Journal of Chemical Physics. 2006;124(4). doi:10.1063/1.2145882.

41. Baumschlager A, Aoki SK, Khammash M. Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. ACS synthetic biology. 2017;6(11):2157–2167.

42. Freddolino PL, Tavazoie S. The dawn of virtual cell biology. Cell. 2012;150(2):248–250. doi:10.1016/j.cell.2012.07.001.

43. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, et al. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biology. 2009;7(7):33–36. doi:10.1371/journal.pbio.1000149.

44. Langan RA, Boyken SE, Ng AH, Samson JA, Dods G, Westbrook AM, et al. De novo design of bioactive protein switches. Nature. 2019;572(7768):205–210. doi:10.1038/s41586-019-1432-8.