Optogenetic Maxwell Demon to Exploit Intrinsic Noise and Control Cell Differentiation Despite Time Delays and Extrinsic Variability

May 4, 2023

M P May[1], B Munsky[1,2]

Abstract

Unnderstanding the effects of noise in control theory is important to () Much work has been done to make control performance robust to noise, but the development of controllers which use noise to their advantage is not fully understood. Motivated by optogentics, we proposed a difficult control problem which requires the exploitation of noise in a stochastic system to break symmetry between two signals. in previous work, we found multiple such controllers which could exploit noise to break symmetry between two cells under a variety of system information. This work extends that analysis to include stochastic systems that trace a moving target, are affected by time delay, intrinsic noise, or measurement uncertainty.

Synthetic biology seeks to create modular components that generate complex and controllable biological behaviors, often at microscopic scales where important regulatory RNA or proteins are present in small numbers. Intrinsic fluctuations at these scales cause biochemical 'noise' that frequently results in undesired neussance effects. However, recent theoretical analyses have shown that noise can sometimes be used to one's advantage when seeking to manipulate nonlinear regulatory systems that would be difficult or impossible to control in a fully deterministic setting. Specifically, it was shown that a noise-exploiting optogenetic controller reminiscent of Maxwell's Demon could systematically force multiple identical systems to reach

different independently-assigned stable points, even for a few special cases where the model is inexact or if observations were incomplete. This work expands upon those analyses to explore in greater detail how control performance is affected by errors or approximations in the model or by temporal delays in the observations. We find that noise-exploiting controllers can remain highly effective despite coarse approximations to the model's scale or incorrect estimations of key model parameters, and these controllers can even retain performance for significant time delays. Together, these findings suggest that noise-exploiting control should be possible in real experiments where models are always approximate, where parameters are always uncertain, and where observations are corrupted by errors.

Keywords: Optogenetic control, Stochastic gene regulation, Synthetic biology

1 Introduction

Synthetic biology seeks to create modular [2] and orthogonal [3] components to sense and actuate [4] complex logical systems [5], and which are capable of performing a variety of advanced biological behaviors [6]. New optogenetic tools have increased our ability to actuate embedded systems within cells reliably and with strong control performance [4, 7–9]. Advances in these two fields have enabled computer programmable control of cellular protein production through the use of external optogenetic inputs and smart microscope techniques [10–12]. These types of digital-synthetic actuators allow for fine-tuned, computer-modulated control of cellular systems that were previously impossible [7, 13] and with fast response times in comparison to chemical diffusion.

Classical and modern control methods like PID control and model predictive control have been implemented in such systems [] which can control synthetic systems to different stable points. Theoretical works have shown that stochastic systems can be controlled to unstable points by oscillating around those points [14] but new control techniques which exploit the full probability distribution information of the system can be used in certain scenarios to control simple systems to unstable points by exploiting the natural noise of protein production to its advantage. Previous control effort focus on deterministic ODE formulations of the control problem, but only a few

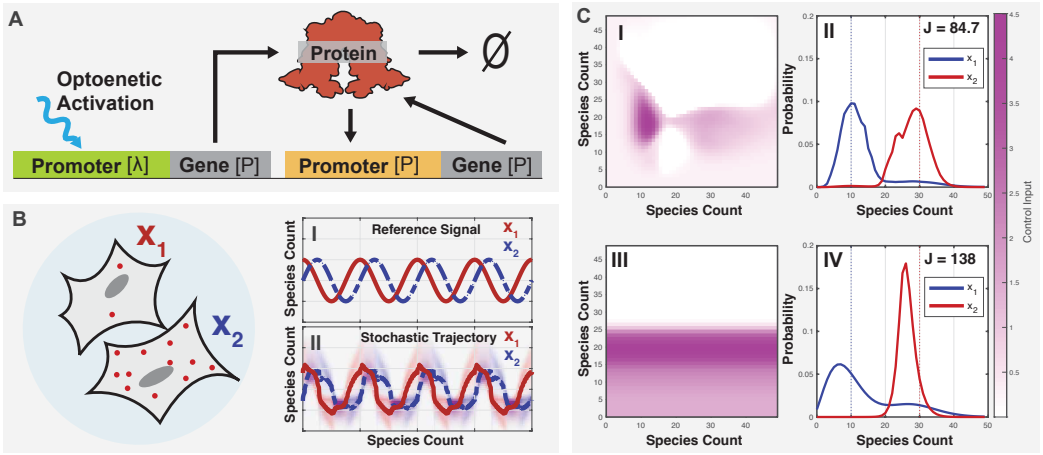

Figure 1: **Optogenetic control of multiple cells using a single light input.** (**A**) Schematic of the light activated genetic system with autoregulation. (**B**) Diagram of the stochastic SIMO control problem using two optogenetic cells sharing an input. (**C**) Adapted from (PAPER). Noise exploiting controllers were optimized to a fully aware control input (I), and partially aware control input(III). Marginal distributions are solved using the finite state projection and show a break in symmetry using both controllers (II and IV)

studies seek to analyze the full chemical master equation to exploit noise to *improve* the control of synthetic biology processes [1, 16].

Szymanska et al. [16] examined a theoretical system inspired by the genetic toggle switch from Kobayashi et al. [**?**], where two gene products repress one another, but where degradation of one protein was under the control of UV radiation via the XXX pathway. Their analysis showed that control was possible despite certain parameter errors and time delays due to maturation of fluorescent proteins and limited observation of the regulatory proteins. To explore this concept in the context of more realistic models and parameters, May et al. [1] fit a simplified stochastic model to reproduce data measured in Baumschlager, et al. [] for the expression from the XXX promoter under the optogenetic control of a UV-activated T7 polymerase (see model in Figure 1A, top promoter). An extended model showed that addition of a positive auto-regulation module (Figure 1A, bottom promoter) could help to maintain an elevated expression phenotype in the presence of UV excitation. Using discrete stochastic models based on direct solutions to the chemical master equation, it was shown that the combination of biochemical noise, nonlinear auto-regulation, and a *single* optogenetic feedback could allow two genetically identical cells with arbitrary initial conditions to be controlled simultaneously to reach different specified phenotypes (Figure 1B). For example, Figure 1C(left) reproduces the heat maps for two control laws, where the level of a single external UV excitation signal was modulated depending on the observed cellular activity, and Figure 1C(right) shows the corresponding marginal stationary distributions that could be achieved under these control laws. Effective control was possible not only for a Fully Aware Feedback Controller (FAC, Figure 1C(top)) that observed the protein counts of all observed cells, but control could be achieved nearly as well for a Partially Aware Feedback Controller (PAC, Figure 1C(bottom)) that observed the protein counts of only one cell. Moreover, it was again seen that these control laws remained effective even in the circumstance where the model used to design the control law was an inexact simplification of the process dynamics to which the controller was later applied.

In this paper, the analysis of this Single-Input-Multiple-Output (SIMO) multicellular control problem is extended to examine how the FAC and PAC controllers are affected by additional model uncertainties, including coarse grained approximations to the system dynamics, uncertainties or errors in system parameters, and delays in the biochemical sensors or optogenetic controllers through which the external feedback is implemented. In the next

section, we briefly describe our methods to formulate the chemical master equation to analyze multiple individual cells in the same spatially homogeneous environment; we define the FAC and PAC controllers and their effects the dynamics of cells within that environment; and we explain how the control law is optimized to achieve desired performance objectives. Next, in the 'Results' section, we explore how granularity of the model formulation, inaccuracy of parametric assumptions, or introduction of time delays would affect the FAC and PAC control performance. Finally, in 'Conclusions' we summarize our findings and discuss the implications that these controllers, and their robustness to modeling errors, may have on future systems in synthetic biology.

2 Methods

In previous work [1], a six-species gene regulation model was developed to describe light-activated association of two T7 split domains (species 1 and 2) to form active T7 polymerases (species 3) which could then in turn associate with inactive T7 promoters (species 4) to create an active gene construct (species 5) that would produce its eventual protein product (species 6). A second, far simpler single-species model was also developed by assuming quasi-steady equilibrium for the first five species. Both models were independently parametrized by fitting to the same experimental data from Baumschlager, et al. [**?**], and then extended to include a secondary self-activated promoter-gene construct for which the expression rate was defined by a Hill function (see Eqn. 1 below). In [1], we showed that a control law identified using the simple model would work to achieve differential control for the more complex model, thus motivating our current study's efforts to explore further how control performance would change upon additional simplifying assumptions or errors in model specification. In the following subsections, we discuss our methods to formulate this model, show how it is analyzed numerically, and discuss the specification of the control objective and the determination of an optimal controller.

2.1 Model

To examine how model approximations would affect the application of noise-enhanced control strategies, we start with the one species model from [1] (see

Table 1: Model Parameters

Parameter	Value	units
κ	0.406	Molecules Per Minute
β	20.0	Molecules
η	8.00	unit less
k_0	0.0001	Molecules per Minute
γ	0.0203	per Minute
u	dynamic	Molecules per Minute

Fig. 1A), which consists of two reactions: protein production and degradation. The nonlinear light-activated protein production rate is given by

$$\nu_1(x,t) = \kappa \frac{x^\eta}{x^\eta + \beta^\eta} + k_0 + u(UV(t)), \tag{1}$$

where x is the instantaneous protein level; κ is the maximum strength of the auto-regulation promoter; β is the concentration at which auto-regulation promoter reaches half its strength; η is the cooperativity in the auto-regulation promoter; k_0 is the promoter leakage rate; and $u(UV(t))$ is the T7 promoter strength. Through external feedback, the level of light excitation, and therefore the T7 promoter strength, $u(UV(t))$, can be modulated in response to the state of the system, thus removing explicit dependence on time. For example, in a system with the observation of two cells, one could define $u(UV(t)) = u^{(\mathrm{FAC})}(x_1, x_2)$ as shown in Fig. 1C(top left) for the fully aware control law, and for a system with the observation of only one cell, one could define $u^{(\mathrm{PAC})}(UV(t)) = u(x_1)$ as shown in Fig. 1C(bottom left) for the partially aware control law. Protein degradation is assumed to be a first order process with rate γ:

$$\nu_2(x) = \gamma x. \tag{2}$$

Parameters describing the auto-regulation promoter, κ, η, β, and k_0, and the degradation rate γ are presented in Table 1 [1].

2.2 Stochastic analyses of the model

To describe the discrete behavior of the above model for a population system with N_c cells, we define the current state of the system as the tuple of the non-negative numbers of proteins in each cell: $\mathbf{X}_i = [x_1, x_2, \ldots, x_{N_c}]_i \in \mathbb{Z}_{\geq 0}$,

where the index i denotes the enumeration of the state within the countably infinite set of all possible states, $\mathbf{X}_i \in \mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots\}$. The stoichiometry vector for a reaction in any given cell is then defined as the change in state following that reaction event (e.g., $\mathbf{X}_i \to \mathbf{X}_i + \mathbf{s}_\mu$), where the $2N_c$ possible reactions are defined in pairs corresponding to production (odd indices) and degradation (even indices) as:

$$\begin{aligned} \mathbf{s}_1 = \mathbf{e}_1,\ \mathbf{s}_2 = -\mathbf{e}_1, \\ \mathbf{s}_3 = \mathbf{e}_2,\ \mathbf{s}_4 = -\mathbf{e}_2, \\ \vdots \\ \mathbf{s}_{2\mathrm{N_c}-1} = \mathbf{e}_{\mathrm{N_c}},\ \mathbf{s}_{2\mathrm{N_c}} = -\mathbf{e}_{\mathrm{N_c}}, \end{aligned} \tag{3}$$

where each $\mathbf{e}_i \in \mathbb{Z}^{\mathrm{N_c}}$ is is a Euclidean vector (i.e., unity for the i^{th} entry and otherwise zero). The corresponding propensity functions are:

$$\begin{aligned} w_1(\mathbf{X}) = u(\mathbf{X},t) + \kappa \frac{x_1^\eta}{x_1^\eta + \beta^\eta} + k_0,\ w_2(\mathbf{X}) = \gamma x_1, \\ w_3(\mathbf{X}) = u(\mathbf{X},t) + \kappa \frac{x_2^\eta}{x_2^\eta + \beta^\eta} + k_0,\ w_4(\mathbf{X}) = \gamma x_2, \\ \vdots \\ w_{2\mathrm{N_c}-1}(\mathbf{X}) = u(\mathbf{X},t) + \kappa \frac{x_N^\eta}{x_N^\eta + \beta^\eta} + k_0,\ w_{2\mathrm{N_c}}(\mathbf{X}) = \gamma x_N. \end{aligned} \tag{4}$$

With these stoichiometry and propensities, we can formulate and run the Gillespie stochastic simulation algorithm [17, 18] to generate representative trajectories of the stochastic process. However, quantifying performance and designing an optimal controller requires a more direct analysis of the chemical master equation (CME). For the same specifications of the stoichiometry and propensity functions, the CME is a linear ODE that describes how probability mass changes in time due to the specified reaction propensities and stoichiometries, which can be written:

$$\frac{d}{dt}P(\mathbf{X}_i) = \sum_{\mu=1}^{2\mathrm{N_c}} \left(-w_\mu(\mathbf{X}_i)P(\mathbf{X}_i) + w_\mu(\mathbf{X}_i - \mathbf{s}_\mu)P(\mathbf{X}_i - \mathbf{s}_\mu)\right). \tag{5}$$

For convenience, the CME can also be formulated more compactly in matrix/vector form as:

$$\frac{d}{dt}\mathbf{P} = (\mathbf{A}_0 + \mathbf{B}\mathbf{u}^{\mathcal{C}})\mathbf{P}, \tag{6}$$

where $\mathbf{P} = [P(\mathbf{X}_1), P(\mathbf{X}_2), \ldots]^T$ is the enumerated probability mass vector for all possible states of the system; $\mathbf{A}_0$ is the infinitesimal generator of the stochastic process due to the autoregulation promoter and degradation events; $\mathbf{u}^{\mathcal{C}} = [u^{\mathcal{C}}(\mathbf{X}_1), u^{\mathcal{C}}(\mathbf{X}_2), \ldots]^T$ is the collection of control inputs associated with each state; and $\mathbf{B}\mathbf{u}^{\mathcal{C}}$ is the contribution that these control inputs make to the infinitesimal generator when included into the feedback process.

More specifically, the open-loop infinitesimal generator, $\mathbf{A}_0$, is constructed according to:

$$[\mathbf{A}_0]_{ij} = \begin{cases} -\sum_{\mu=1}^{2N_c} w_\mu(\mathbf{X}_j), & \text{for } i = j, \\ w_\mu(\mathbf{X}_j), & \text{for } \mathbf{X}_i = \mathbf{X}_j + \mathbf{s}_\mu \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

and the feedback infinitesimal generator, $\mathbf{B}\mathbf{u}^{\mathcal{C}}$, of the controller is constructed according to

$$[\mathbf{B}\mathbf{u}^{\mathcal{C}}]_{ij} = \begin{cases} -N_c u^{\mathcal{C}}(\mathbf{X}_j), & \text{for } i = j \\ u^{\mathcal{C}}(\mathbf{X}_j), & \text{for } \begin{array}{l} \mathbf{X}_i = \mathbf{X}_j + \mathbf{e}_{i_c}, \\ \text{and } i_c = 1, \ldots, N_c \end{array} \\ 0, & \text{otherwise,} \end{cases}, \tag{8}$$

where $u^{\mathcal{C}}(\mathbf{X}_j)$ is the specification of the $\mathcal{C}$ controller in terms of the state, or its partial observations.

For a given controller, the equilibrium distribution of the system ($\mathbf{P}^*$) can be found by solving Eq. 6 and is given by:

$$\mathbf{P}^* = \text{null}(\mathbf{A} + \mathbf{B}\mathbf{u}^{\mathcal{C}}). \tag{9}$$

2.3 Quantification and optimization of control performance

For the single-input-multiple-output control of the multiple cell system, any input signal that is applied to one cell in the population will be felt by all cells in the population. As a result, it is not possible to independently control any individual cell without introducing changes to the behavior of the others. Therefore, an effective controller must be able to strike a balance among the desired behaviors of all cells in the system.

To quantify performance success, we define the performance error score, J, as the expected steady state squared Euclidean distance of the process

from the specified target state, $\mathbf{T}$:

$$J = E\{(\mathbf{X} - \mathbf{T})^2\}. \tag{10}$$

This score is easily calculated by applying a linear operation on steady state probability distribution $\mathbf{P}^*$ from Eq. 9) as follows:

$$\begin{aligned} J &= \lim_{t\to\infty} \mathbb{E}\{|\mathbf{X}(t) - \mathbf{T}|_2^2\}, \\ &= \sum_{i_1,i_2,\ldots} P^*(x_1 = i_1, x_2 = i_2, \ldots) \left[(i - \mathbf{T}_1)^2 + (j - \mathbf{T}_2)^2 + \ldots\right], \\ &= \sum_{i,j,\ldots} P^*_{ij\ldots}(t) C_{ij\ldots} = \mathbf{C}\mathbf{P}^*, \end{aligned} \tag{11}$$

where $\mathbf{C}$ is simply a vector that contains the squared Euclidian distance of each state from the specified target T, i.e., $C_i = |\mathbf{X}_i - \mathbf{T}|_2^2$. As a result of this calculation, J is a non-negative scalar that is zero if any only if $\mathbf{P}^*$ is a delta distribution located exactly at the target vector $\mathbf{T}$.

In this work, we explore two different controller designs that allow for the controller to use two different types of information. A fully aware controller ($\mathbf{u}^{FAC}$) uses direct observation of the protein count of two cells simultaneously while making its control input decision, while the partially aware controller $\mathbf{u}^{PAC}$) uses only the direct observation of a single cell:

$$\mathbf{u}^{FAC} = u(x_1, x_2) \tag{12}$$

$$\mathbf{u}^{PAC} = u(x_1) \tag{13}$$

where x_1 and x_2 are discrete integers greater than or equal to zero that represent the number of proteins in cell one and cell two. In [1], the controllers $\mathbf{u}^{FAC}$ and $\mathbf{u}^{PAC}$ controllers were optimized to minimize J, and both $\mathbf{u}^{FAC}$ and $\mathbf{u}^{PAC}$ were saved as lookup tables after being optimized. In what follows, we will use these controllers exactly and without modification to explore how performance changes when the controllers are applied to systems that are approximations of the model under they were originally designed.

2.4 Scaling for system granularity

One of the major computational obstacles in using the CME or FSP formulations to analyze discrete stochastic chemical kinetics is that the dimension

of the FSP scales linearly with each chemical species, and therefore exponentially with the number of independent species. To circumvent this issue, several previous studies have projected the FSP onto lower dimensional spaces based using various techniques such as time scale separations [**?**], Krylov subspaces [**?**], coarse meshes [**?**, **?**], or principle orthogonal decompositions [**?**]. On the other hand, an experimental obstacle to using CME analyses is that gene expression quantification based on fluorescence intensities (e.g., using fluorescent protein reporters) usually does not provide molecular resolution, but rather measures only relative changes in expression over time.

To address how variations in model or observation scales might impact the effectiveness of noise-based control, we use the controller optimized at one resolution and ask how effective it would be when applied to systems with the same dynamics but with different resolution. To accomplish this rescaling, let M denote some metric for the size of the original system (e.g., the average number of proteins at steady state and maximal UV) and let M' denote the same metric for the system at a different resolution. Toward this goal, we define a granularity parameter ($\alpha = M'/M$) that linearly scales the species populations to increase ($\alpha > 1$) or decrease ($\alpha < 1$) the maximum populations size, while maintaining the dynamics and general behavior of the model. To apply the granularity parameter, we assume that each propensity function, w_μ, from Eqns. 4 is rescaled to a different level of discreteness by substituting

$$w'(\mathbf{X}) = w(\mathbf{X}/\alpha). \tag{14}$$

For example, the production and degradation of protein in cell one would become:

$$w'_1(\mathbf{X}) = u(\mathbf{X}/\alpha, t) + \kappa \frac{(x_1/\alpha)^\eta}{(x_1/\alpha)^\eta + \beta^\eta} + k_0, \;\; w'_2(\mathbf{X}) = \gamma x_1/\alpha. \tag{15}$$

After this rescaling, if $\alpha < 1$ the system dynamics will now occur over a smaller range of protein counts, and if $\alpha > 1$ the system dynamics occurs over a larger range of protein counts.

We note that in transforming the propensity functions, the inputs to the controller have also be scaled by $1/\alpha$ before computing the appropriate functions of that state, e.g., $u^{\mathcal{C}} = u^{\mathcal{C}}(\mathbf{X}/\alpha)$. Because identification of the original control formulation, $\mathbf{u}^{FAC}(x_1, x_2)$ and $\mathbf{u}^{PAC}(x_1)$ only considered integer values for (x_1, x_2), fractional number inputs for after rescaling $(x_1/\alpha, x_2/\alpha)$ are handled by 2D cubic interpolation of the nearest values. Finally, to provide

a consistency relative scoring, the definition of the performance score is also adjusted according to sale magnitudes, i.e.,:

$$\begin{aligned} J' &= \sum_{i=0}^{M'} \sum_{j=0}^{M'} P^*(x_1 = i, x_2 = j)((i/\alpha - \mathbf{T}_1)^2 + (j/\alpha - \mathbf{T}_2)^2), \\ &= \mathbf{C}'\mathbf{P}^* \end{aligned} \tag{16}$$

2.5 Observation and actuation time delays

No real optogenetic control system can provide instantaneous feedback – there will always be delays in biochemical reactions needed for observations (e.g., transcription and translation dynamics as well as folding and maturation of fluorescent proteins), delays in data analysis and decision making (e.g., image processing and calculation of the corresponding control signals), and delays in the actuation dynamics (e.g., activation of optogenetically excitable molecules). To explore the impacts of delays between the time a cellular fluctuation occurs and the time at which the corresponding control action can take effect, we developed a simple time-delay stochastic simulation algorithm. In this analysis, the state history is recorded following every stochastic event, allowing for reconstruction of the piecewise constant history of the species' populations, e.g., $x_1(s)$ and $x_2(s)$ for all $s \in [0, t]$. Using this information, the time-delayed control signal at time t can be specified as:

$$u_\tau(t) = \begin{cases} 0, & \text{for } t \leq \tau, \\ u^{\mathcal{C}}(x_1(t-\tau), x_2(t-\tau)), & \text{for } t > \tau, \end{cases} \tag{17}$$

where τ is the time delay between observation and actuation, and $u^{\mathcal{C}}$ is the previously optimized control law (e.g., $\mathbf{u}^{\text{FAC}}$ or $\mathbf{u}^{\text{PAC}}$). We note that the time delay stochastic process was only simulated using the SSA because to our knowledge an appropriate direct FSP/CME integration procedure has not yet been developed.

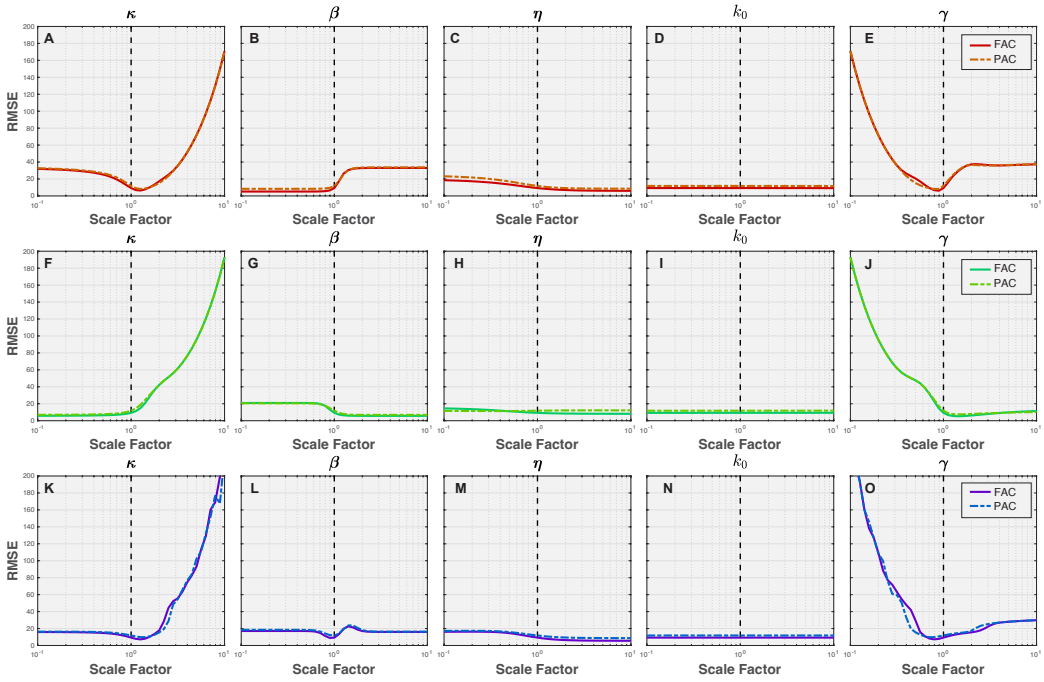

Figure 2: Parameter sweeps using the FAC and PAC show a broad range of control performance in cell 1 (**A**), in cell 2 (**B**), and in both cells (**C**).

3 Results

3.1 Effects of Parameter Errors or Extrinsic Uncertainties

Exact parameters values are typically unknown for real stochastic process, and even genetically identical single-cell systems can exhibit cell-to-cell heterogeneity in parameters due to extrinsic variations outside of the specific pathways of interest. Estimates of parameters can be taken, but not exact measurement. These parameter uncertainties could influence system performance. To explore the sensitivity of control performance to unknown errors in the model parameters, sensitivity analysis was performed for each model parameter and for the situation where the corresponding parameter was modified in either one cell at a time (i.e., to simulate extrinsic noise) or in both cells simultaneously (i.e., to simulate parameter errors). For each perturbation analysis, both the $\mathbf{u}^{\mathrm{FAC}}$ and $\mathbf{u}^{\mathrm{PAC}}$ controller were analyzed over a range of parameters from one-half to two-fold of the original parameters. Figure 2A shows the results for these parameter sweeps when a single parameter in Cell 1 is change while the parameters of Cell 2 is held fixed, and Fig. 2B shows the opposite when changes Cell 1 has fixed parameters and Cell 2 is modified. Finally, 2B shows the performance change when the corresponding parameter is changed in both Cell 1 and Cell 2 at the same time.

Modification to parameters was found to produce a broad range of effects, where some changes increase performance, some decrease performance, and others have little or no effect. For example, increasing β in Cell 1 worsens performance while increasing β in Cell 2 improves performance (compare second column in Fig. 2A and 2B). In some cases, these effects were not monotonic; for example, increasing κ in Cell 1 (Fig. 2A, left column) would be highly advantageous up to a limit after which the control performance degrades rapidly. In other cases (such as for k_0), the effect of parameter perturbations on performance is insignificant even for relatively large (s=2) perturbations.Similarly, Fig. 2C shows that when parameters of both Cell 1 and Cell 2 are jointly changed, these changes could also improve or detract from control performance.

When examining the effects of changing parameters when using the $\mathbf{u}^{\mathrm{FAC}}$ (solid lines) or the $\mathbf{u}^{\mathrm{PAC}}$ (dashed lines), general trends typically remained the same in that changes to a parameter which cause a decrease in the control performance of the FAC also tend to decrease the performance of the PAC.

Some parameters (e.g., κ, γ) reached minima that performed better than the original, which suggests that the physical system itself has room for optimization beyond the original design that could lead to better control performance, *even without changing the control law*. One explanation for why changes to k_0 had no effect on control performance is that the original parameter was so small that it may require a larger scaling on a different magnitude before changes are observed. In a few circumstances, changes in the parameters κ or γ led to the surprising situation where the PAC controller, which only observes one cell at a time, was more robust to parameter errors than was the FAC controller, which can observe both cells, even to the extent that the PAC controller could sometimes outperform the FAC. We stress that this effect is due to the mismatch between the assumed and the actual parameters, and re-optimizing the two controllers for the new parameter combination will always yield better control results for the FAC compared to the PAC.

To assess the potential robustness of noise-enhanced control of the proposed synthetic auto-repression toggle switch, we adopted the model from [1] as well as the fully aware and partially aware controllers, $u^{\text{FAC}}(x_1, x_2)$ and $u^{\text{FAC}}(x_1)$, respectively as described in the methods section. In the following subsections, we explore how well these controllers would perform in the more realistic setting where the approximate model used to define the controller no longer matches exactly to the dynamics of the process that the controller is being used to manipulate.

We first explore how the model defined for a given assumed system size (i.e., in which protein numbers have a range of M) when applied to a system with larger or smaller systems sizes (i.e., where protein numbers now have a range of $M' = \alpha M$. This modification could arise from two situations related to numerical convenience or inexact resolution in the measurement of concentrations. First, from the perspective of numerical convenience when optimizing a controller, it can be advantageous to reduce the assumed range of protein levels in order to reduce the dimension of the CME and FSP analyses, thereby achieving a more tractable optimization problem. For example, the current model (see Methods) takes 19.3 seconds to solve the FSP analysis when $\alpha = 1$ but the same analysis takes 24619 seconds to solve when $\alpha = 10$. Since numerical optimization of the controller can take many millions of calculation for different control laws, learning a controller using $\alpha = 1$ that works for $\alpha = 10$ would yield substantial savings in computational effort. The second situation under which sensitivity to system scaling could become important is when measurements are relative and do not provide

single-molecule resolution. For example, in measurements like those on [**?**] which are based on fluorescence protein intensities, it is not clear how many proteins correspond to what level of fluorescence intensity, and the scalar α is an unknown quantity that must be estimated (e.g., through calibration with single-molecule measurements or dilution experiment [**?**]). It is well established that the relative size of stochastic fluctuations (i.e., noise) in a chemical process decreases with the inverse square root of the process scale [**?**]. As the system size increases (with all species being scaled by the same factor), the dynamics approach those of a deterministic process with probability one (with zero-measure exceptions only for initial conditions lying on manifolds that separate attractors for different steady state behaviors) [**?**]. For the purpose of noise-induced control, this introduces a tradeoff where noise is critical to break symmetry and enable differential control, yet noise is also deleterious to maintaining desired states once they have been achieved. Considering these competing objectives, it is not immediately clear how the reduction of noise through the addition of system granularity will affect control performance.

To shed light on this tradeoff, control performance scores for the $\mathbf{u}^{\text{FAC}}$ and $\mathbf{u}^{\text{PAC}}$ controllers were calculated at different levels of granularity (α) between 0.1 and 2.0. Figure 3 (Ai - Fi) shows the joint (left plots) and marginal (right plots) distributions of the two cells relative to the specified target position (circles in joint distributions and vertical dashed lines in marginal distributions) for the FAC controller and for α=.2 (A,B), 1 (C,D) and 2 (E,F). As the granularity changes from α=0.2 , to α=2, the distributions become focused at a point near the target, and the performance score improves from J =333 to 85 to 27, respectively. The PAC controller performance also improves considerably with alpha as shown in Fig. 3 (G-L).

3.2 Effects of changes to system granularity

Figure 3M shows the trend of the performance score versus alpha for both the FAC (solid cyan line) and the PAC (solid magenta line) controller. This improvement in performance appears to approach a small value as the granularity goes to infinity, but since the size of the FSP increases with the square of the system size, systems much larger than α=2 (where $\mathbf{A}_0 \in \mathbb{R}^{10^4 \times 10^4}$) become more difficult to calculate. To bypass this limit in the FSP, sixteen SSA simulations were used to sample the CME of a system with a much larger volume of α=100. Each SSA was ran for 5×10^7 minutes and only the

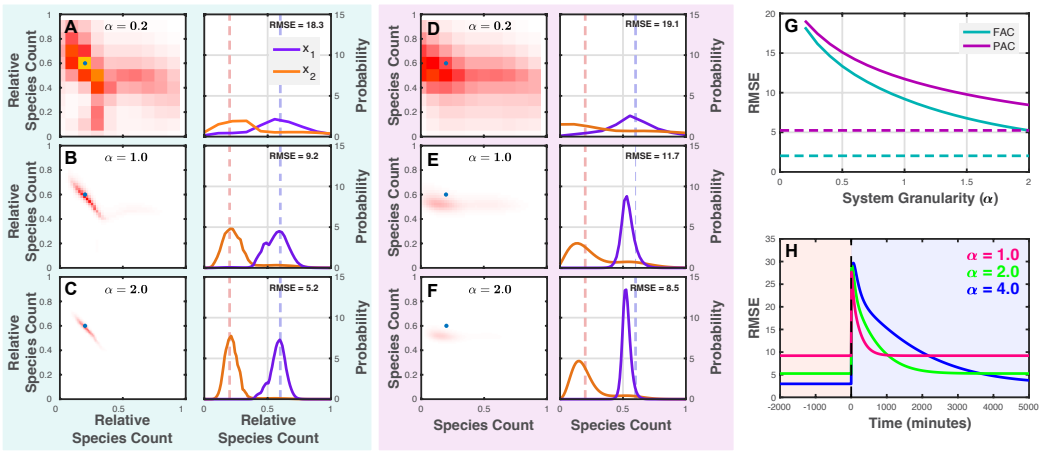

Figure 3: Systems with high granularity were shown to have increased control performance but slower response times.

last 4×10^7 minutes were sampled to estimate the stationary distribution and the calculate the performance score. The performance score estimates of this high granularity SSA using the FAC and PAC were 4.13 and 27.5 respectively, which are plotted as dashed lines in Fig. 3G. Although it is unclear if further performance improvements could be obtained with further increases to the system scale, for all cases considered so far, we found that both controllers monotonically improved with increased α and that $\mathbf{u}^{\text{FAC}}$ always outperforms the $\mathbf{u}^{\text{PAC}}$.

We note however, that the improved steady state performance for systems with larger scales, comes at a cost that they also take longer to reach a steady state distribution. This tradeoff is illustrated in Fig. **??** which shows the performance score $J(t)$ as the system adapts to a change from the original target point $\mathbf{T} = [30, 10]$ to its mirror point $\mathbf{T}_m = [10, 30]$. One hundred SSA simulations of the flip-mirroring process were simulated while measuring the average score all simulations over time. Fig. **??** (A and B) shows the average score and protein counts of the simulation when $\alpha = 1$, Fig. **??** (C,D) when $\alpha = 10$, and Fig. **??** (E,F) when $\alpha = 100$.

From the figures it is clear that the smaller α creates faster dynamic but also have higher scores due to inherit noise. When $\alpha = 10$, the system becomes much less noisy and have lower scores but takes some time to respond. When α is too large at 10, the system contains very small levels of noise, but

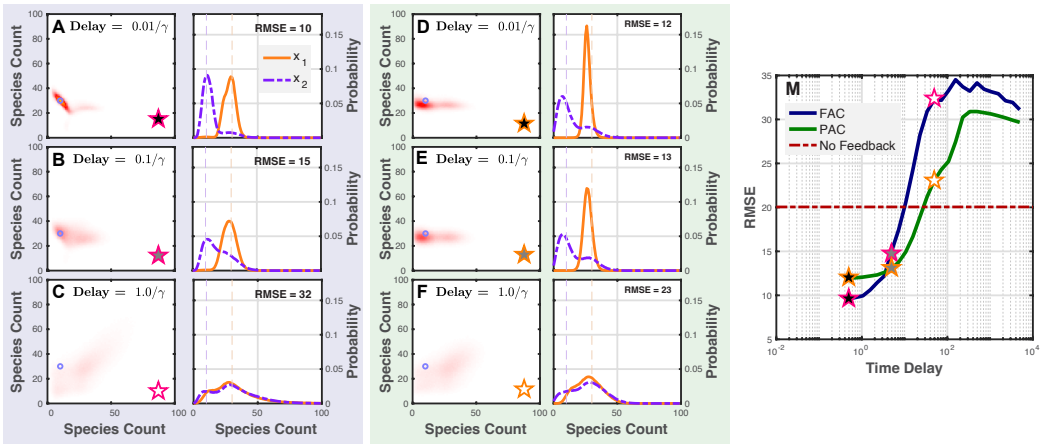

Figure 4: Stochastic simulations of the time delayed system show decreasing control performance using either controller.

the dynamics

3.3 Effects of Time Delays on Controller Performance

Feedback control can only be effective if one can quickly make measurements, compute adjustment to the control signal, and implement the needed change to the system. As the time required for any of these steps increases, control performance will be degraded, perhaps even leading to large fluctuations or instability. To explore how time delays would affect the noise-enhance controller $\mathbf{u}^{\mathrm{FAC}}$ and $\mathbf{u}^{\mathrm{PAC}}$, we generated large sets of time-delayed stochastic simulations (see Methods) for different lengths of the time delay. Each SSA was sub-sampled for 1000 times over 10000 minutes of simulation time after a burnin period of 10000 minutes.

Figure 4 shows the joint and marginal distribution of each controller and time-delay pair and the resulting score, with panels A-F showing results for the FAC controller and panels G-L showing results for the PAC controller. Figure 4M summarizes these results by plotting the score of both controllers versus the time delay. From the figures, it is clear that performance is rapidly degraded as the delay approaches and then exceeds the characteristic time scale of the process. At very small time delays (below $\tau = 0.07/\gamma = 0.0142$ min), the FAC outperforms the PAC (J =87 versus 146 at $\tau = 0.01/\gamma = 0.00203$ min) but at moderate time delays (above $\tau = 0.07/\gamma = 0.0142$ min)

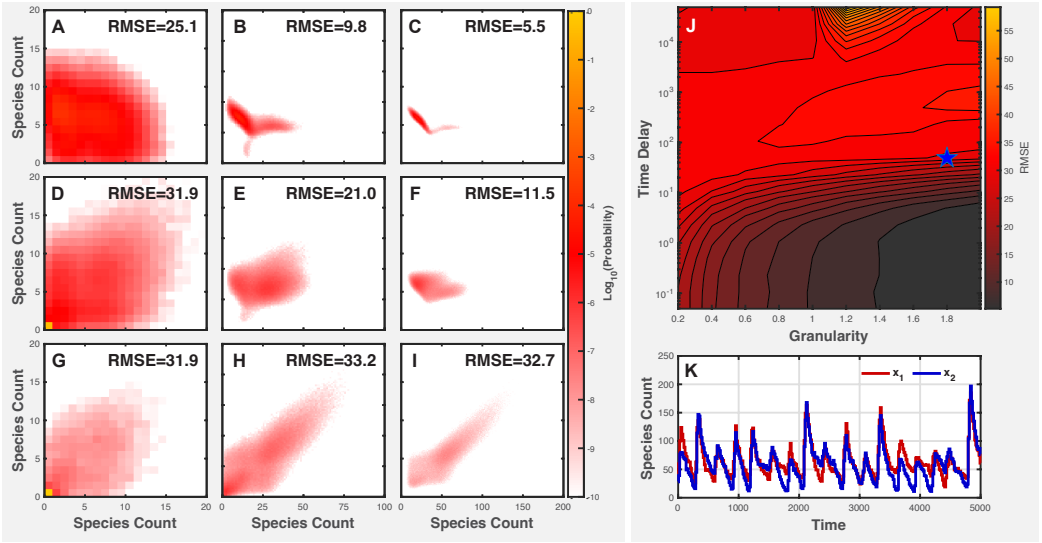

Figure 5: Stochastic simulations of the FAC driven system (A) and PAC driven system (B) show different failure modes at at different (τ, α) pairs (C).

the PAC outperforms the FAC (J =170 versus 219 at $\tau = 0.1/\gamma = 0.0203$ min).

This data taken together show that time much larger that $0.07/\gamma = 0.0142min$ is detrimental to the FAC controller; delays beyond $0.2/\gamma = 0.0406min$ are detrimental for the PAC controller; and the best choice in controller is dependent the level of time delay in the system. For extreme levels of time delay, both systems lose their asymmetry, and their scores become much worse (J =1041 and 931). A no-feedback controller was developed which required no inputs. Such a controller is immune to time delay, and received a score $J = 402$. These data show that at large time delays above , attempting to improve control with feedback is worse than no-feedback.

Joint analysis of time delay and granularity together was performed to better understand if granularity can undo the effect of time delay. System control performance was analyzed over a two dimensional domain of points (τ, α) using sixteen stochastic simulations for XXX minutes. These simulations were driven using the FAC controller Fig 5 (A), and PAC controller Fig 5 (B). The landscape of the FAC and PAC control performance Fig 6(A and B)over the domain shows a valley at low τ, and a plateau at large τ. This

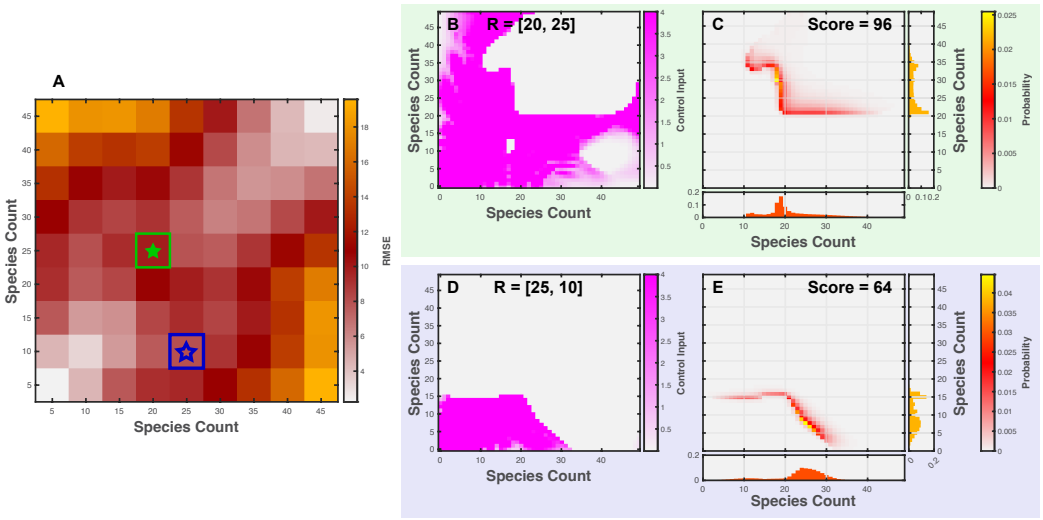

Figure 6: Controllers optimized to a 2d domain of target points (A) show good control performance in the central region despite a broad range of targets (B and C).

suggests that increasing system scale at large time delays is not beneficial to control performance. To better understand the appearance of the landscape, trajectories in different regions were compared. Trajectories at the edge of the plateau region (blue star) showed bursty behavior Fig 6(C)(blue lines), while trajectories at the edge of the valley Fig 6(C) (red lines) showed oscillatory behavior. Despite the small different in time delay between the red and blue trajectory, a stark change to the overall behavior was observed.

Many controllers were optimized to different target points to determine if certain reference points were harder to control than others. Controllers were optimized over a two dimensional domain of target points (T_1, T_2)between between five and fourty-five. Figure 6 (A) shows the control performance of the system after optimizing each controller to each point in the domain. Figure 6 (B and C) show the optimized control input and the steady state of two points in the system. The unstable point in the rate equation matches the unstable point near $(20, 20)$ in this domain.

Dynamic reference targeting was performed by cycling though individual controllers which were optimized to a single target point at a fixed frequency. Thirty-two controllers were optimized along a pathway in order to target an in-sync reference point 7(Synchronous), a phase lagged reference point

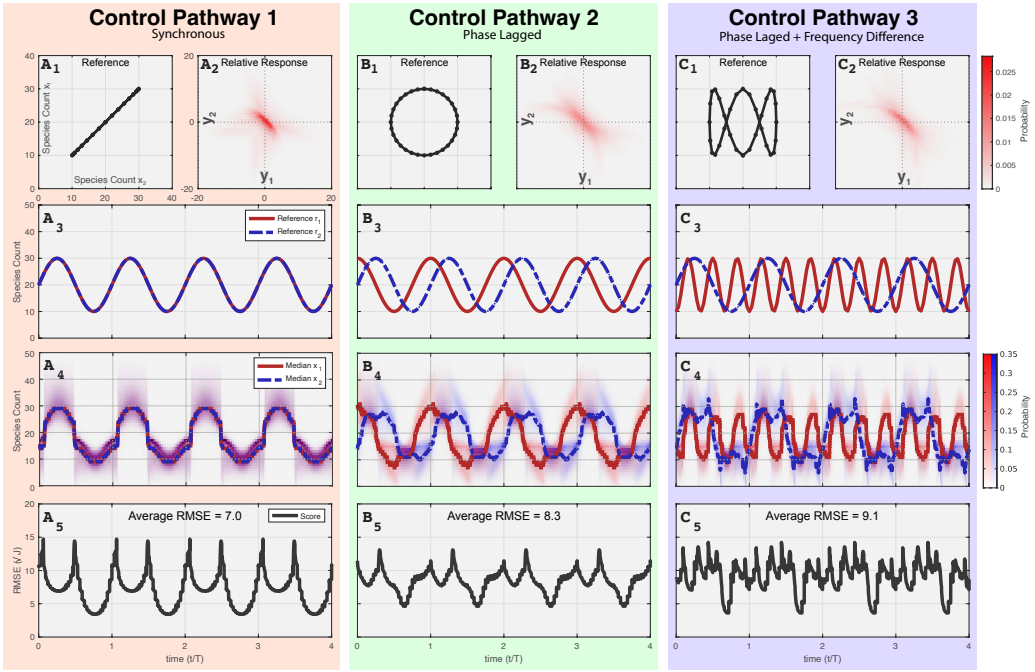

Figure 7: Finite state projections of the slowly driven system demonstrate good control performance under a variety of phase and frequency scenario.

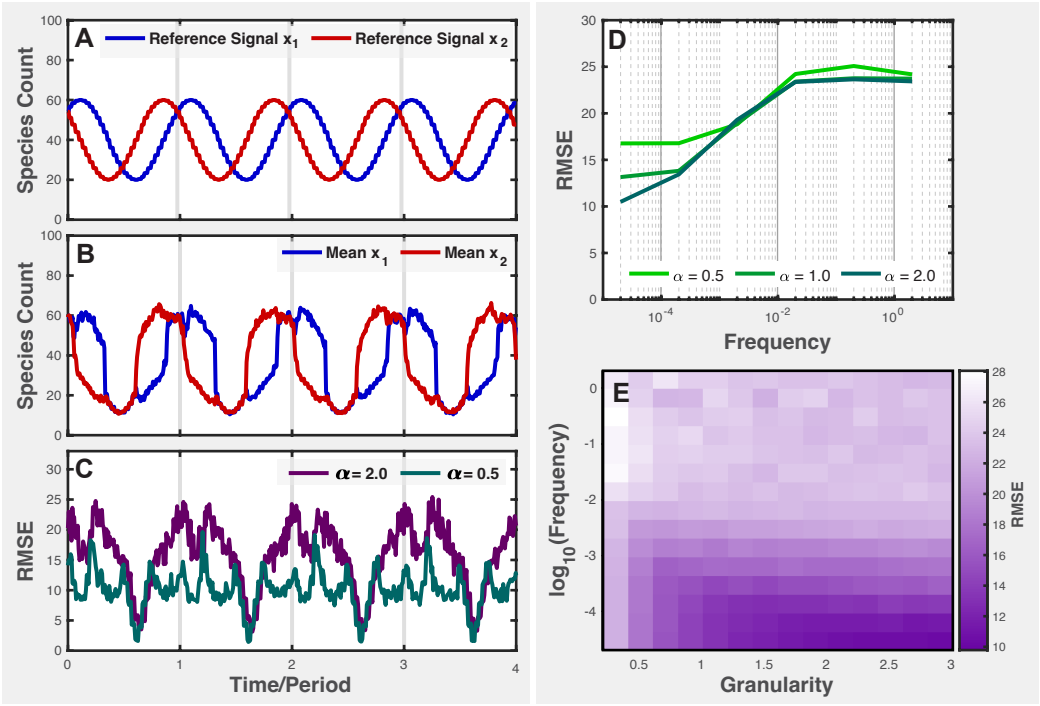

Figure 8: Stochastic simulations driven using a phased lagged controller show improved score at high granularity and low frequency.

7(Phase Lagged), or frequency separated reference point7(Phase Lagged and Frequency). The in phase signal is given by $r(t) = XXX$, the phase-lagged signal is given by $r(t) = XXX$, and the frequency driven signal is given by $r(t) = XXX$ FSP simulations of cyclically driven systems were slowly driven at a frequency of 1/5000 seconds and their medians are shown in the Fig 7(A4, B4,C4)(red cloud, blue cloud). Root mean squared error of each simulation are shown in Fig 7(A5, B5,C5). Simulations show that in-phase control performed the best with an RSME of 7.0 despite moving over the unstable point in the system. When the system is driven with a phase-lag the RSME of the score increases to 8.3 and when driven at a different frequency the RMSE goes up to 9.1.

Although the control of a dynamic reference point using a Piecewise controller is possible, the effect of cycling frequency is also important.

16 SSA simulations of the phase-separated driving controller were performed on a system over a two dimensional domain of points (α, f) for a

time of XXX minutes. Comparisons of control performance at $\alpha = 0.5$ and $\alpha = 2.0$ (Fig (C))show that control performance improves with system scale. 16 SSA simulations of the high granularity system driven by phase-lag were performed over a range of frequency values. These analysis show that at low frequency the systems have a noise floor which decreases with increasing alpha, but was the system were driven rapidly, all systems became equally uncontrollable. At high frequency, the scores never got worse than $J = 650$. At high frequency the control performance plateaus, while at lower frequency the system improves with increasing α Fig(E).

4 Conclusion

Key results from [1] showed that noise was a key requirement for the development of a noise-exploiting controller, and that deterministic systems could not be controlled to two different stable points if they both started at the same initial condition. As a stochastic system becomes increasingly more granular, it also becomes less noisy and more similar to its ODE solution which is known to be impossible to control to two different fates, but here, we have shown that the removal of noise through system granularity led to better control performance. Therefore, stochastic controllers may perform work best on models which are nearly deterministic, but not quite. Unfortunately, such systems were found to take longer time to achieve steady state.

The ability to analyze the model at one system granularity apply it elsewhere might help analyze models for which the computational effort may be too large. Since the computation time of the FSP solution to the CME grows with the square of the number of states, this can cause an explosion in computational requirements for large systems. One alternative is to learn a controller at a computationally feasible number of states and apply them to large systems which cannot be solves for using the FSP.

Parameter perturbation analysis showed that there is room to improve control performance by adjusting system parameters, and that joint optimization of the controller with the system parameters may lead to better control performance. It also shows that populations of cells where each cell might have its own parameter combination could still be reasonably controlled for small changes in parameter scale, but large changes could be detrimental.

Time delay analysis showed that increasing time delay decreased control

performance. It was also found that a controller with less control information outperformed a controller with more information at higher levels of time delay. We believe this is happening because a controller with more information can afford to be more aggressive to implement its control, and time delay can cause this aggression to backfire. A controller was optimized which required no information in May et al which had a score of 402. This controller should be immune to time delay since no feedback is required for this controller. Here we saw both controllers did performed worse than 402 at very large time delay.

References

[1] M. May and B. Munsky, "Exploiting noise, nonlinearity, and feedback to differentially control multiple synthetic cells with a single optogenetic input," pp. 1–28, 2021. eps

[2] A. H. Ng, T. H. Nguyen, M. Gómez-Schiavon, G. Dods, R. A. Langan, S. E. Boyken, J. A. Samson, L. M. Waldburger, J. E. Dueber, D. Baker, and H. El-Samad, "Modular and tunable biological feedback control using a de novo protein switch," pp. 265–269, 2019.

[3] C. C. Liu, M. C. Jewett, J. W. Chin, and C. A. Voigt, "Toward an orthogonal central dogma," *Nature Chemical Biology*, vol. 14, no. 2, pp. 103–106, 2018.

[4] M. B. Sheets, W. W. Wong, and M. J. Dunlop, "Light-Inducible Recombinases for Bacterial Optogenetics," *ACS Synthetic Biology*, vol. 9, no. 2, pp. 227–235, 2020.

[5] T. M. Groseclose, R. E. Rondon, Z. D. Herde, C. A. Aldrete, and C. J. Wilson, "Engineered systems of inducible anti-repressors for the next generation of biological programming," *Nature Communications*, vol. 11, no. 1, pp. 1–15, 2020. [Online]. Available: http://dx.doi.org/10.1038/s41467-020-18302-1

[6] J. Shin, S. Zhang, B. S. Der, A. A. Nielsen, and C. A. Voigt, " Programming Escherichia coli to function as a digital display ," *Molecular Systems Biology*, vol. 16, no. 3, pp. 1–12, 2020.

[7] A. Baumschlager, S. K. Aoki, and M. Khammash, "Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control," *ACS synthetic biology*, vol. 6, no. 11, pp. 2157–2167, 2017.

[8] S. Y. Chen, L. C. Osimiri, M. Chevalier, L. J. Bugaj, T. H. Nguyen, R. A. Greenstein, A. H. Ng, J. Stewart-Ornstein, L. T. Neves, and H. El-Samad, "Optogenetic Control Reveals Differential Promoter Interpretation of Transcription Factor Nuclear Translocation Dynamics," *Cell Systems*, vol. 11, no. 4, pp. 336—-353.e24, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.cels.2020.08.009

[9] G. Lillacci, Y. Benenson, and M. Khammash, "Synthetic control systems for high performance gene expression in mammalian cells," *Nucleic acids research*, vol. 46, no. 18, pp. 9855–9863, 2018.

[10] Z. R. Fox, S. Fletcher, A. Fraisse, C. Aditya, and S. Sosa, "MicroMator: Open and Flexible Software for Reactive Microscopy," *bioRxiv*, pp. 1–9, 2021. [Online]. Available: http://biorxiv.org/cgi/content/short/2021.03.12.435206v1

[11] A. Baumschlager and M. Khammash, "Synthetic Biological Approaches for Optogenetics and Tools for Transcriptional Light-Control in Bacteria," *Advanced Biology*, vol. 5, no. 5, p. 2000256, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adbi.202000256

[12] J. B. Lugagne, S. Sosa Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen, "Balancing a genetic toggle switch by real-time feedback control and periodic forcing," *Nature Communications*, vol. 8, no. 1, pp. 1–7, 2017. [Online]. Available: http://dx.doi.org/10.1038/s41467-017-01498-0

[13] M. Rullan, D. Benzinger, G. W. Schmidt, A. Milias-Argeitis, and M. Khammash, "An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation," *Molecular Cell*, vol. 70, no. 4, pp. 745–756.e6, 2018. [Online]. Available: https://doi.org/10.1016/j.molcel.2018.04.012

[14] A. Guarino, D. Fiore, D. Salzano, and M. Di Bernardo, "Balancing Cell Populations Endowed with a Synthetic Toggle Switch via Adaptive

Pulsatile Feedback Control," *ACS Synthetic Biology*, vol. 9, no. 4, pp. 793–803, 2020.

[15] B. Munsky, G. Neuert, and A. Van Oudenaarden, "Using gene expression noise to understand gene regulation," *Science*, vol. 336, no. 6078, pp. 183–187, 2012.

[16] P. Szymańska, N. Gritti, J. M. Keegstra, M. Soltani, and B. Munsky, "Using noise to control heterogeneity of isogenic populations in homogenous environments," *Physical Biology*, vol. 12, no. 4, 2015.

[17] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 1-3, pp. 404–425, 1992.

[18] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.