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

June 26, 2023

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

Abstract

Previous research has focused on robust control performance in the presence of noise, but understanding controllers that exploit noise remains incomplete. Motivated by Maxwells-Demon, we previously proposed a difficult control problem which requires the exploitation of noise in a stochastic system to break symmetry between two signals. 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. 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 where models are always approximate, where parameters are always uncertain, and where observations are corrupted by errors.

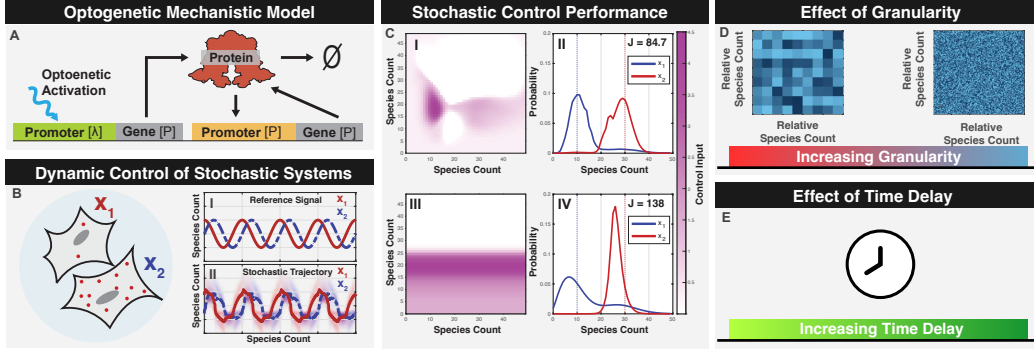Keywords: Stochastic Control, Gene Regulation, Optogenetics

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)

1 Introduction

Noise is a common theme throughout many areas of science, including control theory XXX, gene regulation XXX, mRNA dynamics XXX, and synthetic biology XXX. Excellent work has been done to enhance the robustness of control processes against noise. However, these systems are often controlled in spite of noise rather than working with it. Many stochastic systems can be modeled using a noise-free approach with additive noise but, a detailed analysis using the full master equation could uncover interesting properties, potentially leading to novel control mechanisms.

The cell's drive towards homeostasis requires dealing with the inherently chaotic and noisy processes that reside within it. Brownian motion, stochastic gene regulation, and mRNA dynamics collectively generate a fundamentally noisy environment that cells must effectively manage. Despite these challenges, cells generally demonstrate strong capability to control these noisy processes. However, stochastic control theory is not well-developed, despite having substantial knowledge on how to model them.

Synthetic biology aims to develop modular [2] and orthogonal [3] components to sense and manipulate [4] complex logical systems, enabling them

2

to exhibit a wide range of advanced biological behaviors [6]. Advancements in optogenetics have enhanced our ability to reliably actuate embedded systems within cells, offering precise control [4, 7–9]. These developments have facilitated computer-programmable regulation of cellular protein production through external optogenetic inputs and smart microscope techniques [10–12]. These digital-synthetic actuators enable fine-tuned, computer-modulated control of cellular systems, previously unattainable, with faster response times compared to chemical diffusion [7, 13] . 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 guided towards unstable points by oscillating around those points [14]. However, new control techniques that leverage the complete probability distribution information of the system can exploit the inherent noise of protein production to achieve better control. Previous control efforts primarily focusted on deterministic ordinary differential equation (ODE) formulations of the control problem, only a limited number of studies have delved into analyzing the full chemical master equation to harness noise to improve the control of biological processes [1, 16].

inspired by the genetic toggle switch from Kobayashi et al. [**?**], Szymanska et al. [16] showed that stochastic 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.

This paper extends the analysis of the Single-Input-Multiple-Output (SIMO) multicellular control problem by examining the impact of model uncertainties on FAC and PAC controllers. These uncertainties include coarse-grained system dynamics, parameter errors, and controller delays. Our methods involve formulating the chemical master equation to analyze cells in a state space, defining the controllers and their effects on cell dynamics, and optimizing the control input for desired performance. In the 'Results' section, we explore the effects of model granularity Figure 1D parameter inaccuracies,

3

and time delays Figure 1E on FAC and PAC performance and add a simple scheme for controlling cells to dynamically changing point. Using discrete stochastic models based on the chemical master equation, we demonstrate that combining biochemical noise, nonlinear auto-regulation, and a single optogenetic feedback could control two genetically identical cells with different initial conditions to reach different desired signals at different frequency and phase (Figure 1B).

Figure 1C(left) presents heat maps for two control laws, which a single external UV excitation signal based on observed cellular state. Figure 1C(right) shows the corresponding achievable marginal stationary distributions. Effective control was possible with both Fully Aware Feedback Controllers (FAC, Figure 1C(top)) and Partially Aware Feedback Controllers (PAC, Figure 1C(bottom)), even when the control model was an approximate simplification of the actual process dynamics.

2   Methods

A gene regulation model consisting of six species was developed by May et al. [1]. It described the light-activated association of two T7 split domains (species 1 and 2) to form an active T7 polymerase (species 3). These active polymerase could then associate with inactive T7 promoters (species 4), resulting in the formation of an active gene (species 5) that produced the desired protein product (species 6).

Additionally, a simple single-species model was developed, assuming quasi-steady equilibrium for the first five species. Both models were independently parameterized using experimental data from Baumschlager et al. [?]. Furthermore, an extension was made to incorporate a secondary self-activated promoter-gene construct, where the expression rate was determined by a Hill function 1.

2.1   Model

To assess the impact of model approximations on the implementation of noise-enhanced control strategies, we begin with the one-species model proposed by May et al. [1] (refer to Fig. 1A). This model comprises of two reactions: nonlinear protein production and degradation and it is defined by

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

the following equation:

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

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. Feedback enables the external modulation of the light input using the state of the system, thereby controlling the T7 promoter strength as a function of state rather than time and eliminating the explicit time dependance in $u(UV(t))$.

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

5

reactions are defined in pairs corresponding to production (odd indices) and degradation (even indices) as:

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

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

$$w_1(\mathbf{X}) = u(\mathbf{X}, t) + \kappa \frac{x_1^\eta}{x_1^\eta + \beta^\eta} + k_0, \ w_2(\mathbf{X}) = \gamma x_1,$$

$$w_3(\mathbf{X}) = u(\mathbf{X}, t) + \kappa \frac{x_2^\eta}{x_2^\eta + \beta^\eta} + k_0, \ w_4(\mathbf{X}) = \gamma x_2,$$

$$\vdots$$

$$w_{2N_c-1}(\mathbf{X}) = u(\mathbf{X}, t) + \kappa \frac{x_N^\eta}{x_N^\eta + \beta^\eta} + k_0, \ w_{2N_c}(\mathbf{X}) = \gamma x_N. \tag{4}$$

The stoichiometry and propensities allow us to use the Gillespie stochastic simulation algorithm [17, 18] to generate representative trajectories of the stochastic process. However, a direct analysis of the chemical master equation (CME) is necessary for performance quantification and optimal controller design. The CME, based on the same stoichiometry and propensity functions, is a linear ODE that describes the time-dependent changes in probability mass due to the specified reaction propensities and stoichiometries. It can be expressed as:

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

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

6

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 $\mathbf{Bu}^{\mathcal{C}}$ 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, $\mathbf{Bu}^{\mathcal{C}}$, of the controller is constructed according to

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

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:

$$\mathbf{P}^* = \mathrm{null}(\mathbf{A} + \mathbf{Bu}^{\mathcal{C}}). \tag{9}$$

2.3 Quantification and optimization of control performance

In the SIMO control of the multiple cell process, any input signal applied to one cell affects all cells in the population. Thus, it is not possible to independently control individual cells without impacting others. An effective controller must 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, $\mathbf{T}$:

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

7

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

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

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 $\mathbf{T}$.

We studied two controller designs: the fully aware controller ($\mathbf{u}^{FAC}$) that utilizes simultaneous protein count observations from both cells, and the partially aware controller ($\mathbf{u}^{PAC}$) that relies on observations from a single cell.

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

$$\mathbf{u}^{PAC} = u(x_2) \tag{13}$$

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], we optimized the controllers $\mathbf{u}^{FAC}$ and $\mathbf{u}^{PAC}$ to minimize J and saved them as lookup tables.

2.4   Scaling for system granularity

Previous studies have addressed the computational challenge of analyzing discrete stochastic chemical kinetics by projecting the FSP onto lower-dimensional spaces using techniques such as time scale separations [?], Krylov subspaces [?], coarse meshes [?,?], or principle orthogonal decompositions [?]. However, using CME analyses is often hindered by an experimental limitations. Gene expression quantification based on fluorescence intensities, such as using fluorescent protein reporters, typically captures relative changes in expression over time rather than providing direct molecular counts.

It would be convenient to create a parameter that could increase or decrease the intrinsic noise of a system without changing the overall dynamics of the system. This would enable the analysis of models with more or less

noise without the need of new unique models each time. A rescaling of the system could decrease the intrinsic noise while maintaining the same general behavior. To accomplish this rescaling, let M denote the size of the origional unscaled system and let M' denote the size of the newly scaled system.

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:

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

We note that in transforming the propensity functions, the inputs to the controller must 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

Delays are inherent in optogenetic control systems due to biochemical reactions involved in observations, data analysis, decision making, and actuation dynamics. To investigate the effects of time delay on control actuation, we devised a time-delay stochastic simulation algorithm. This algorithm records

the state history after each stochastic jump, enabling reconstruction of the population history of the species and control input. 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.

2.6   Dynamic Setpoint Control

Although the control to a static set point is at the limits of currently feasible calculations, the extension to dynamically moving set point is much harder and requires untractible numerical simulation or further mathematical exploration. These solutions require numerical simulations which scale poorly with the number of states and quickly become impossible to perform. Therefore, a simple alternative is to sweep over a set of steady state controllers in Piecewise fashion that mimics a dynamically moving path.

We developed a dynamic set point controller which periodically sweeps over a predefined path in state space in a discrete number of steps. Let the path $\mathbf{r}$ be given by

$$\mathbf{r}(\mathbf{x}) = [\mathbf{r_1}(\mathbf{x_n}), \mathbf{r_2}(\mathbf{x} - \mathbf{N})] \tag{18}$$

where $x_n \in [0, n]$, r_1 and r_1 an arbitrary periodic function that describes the desired target state for x_1 and x_2. Then x_n can be broken down into n steps

$$x_n(t) = mod(n * t/T, n) \tag{19}$$

And the path can be walked over in time using

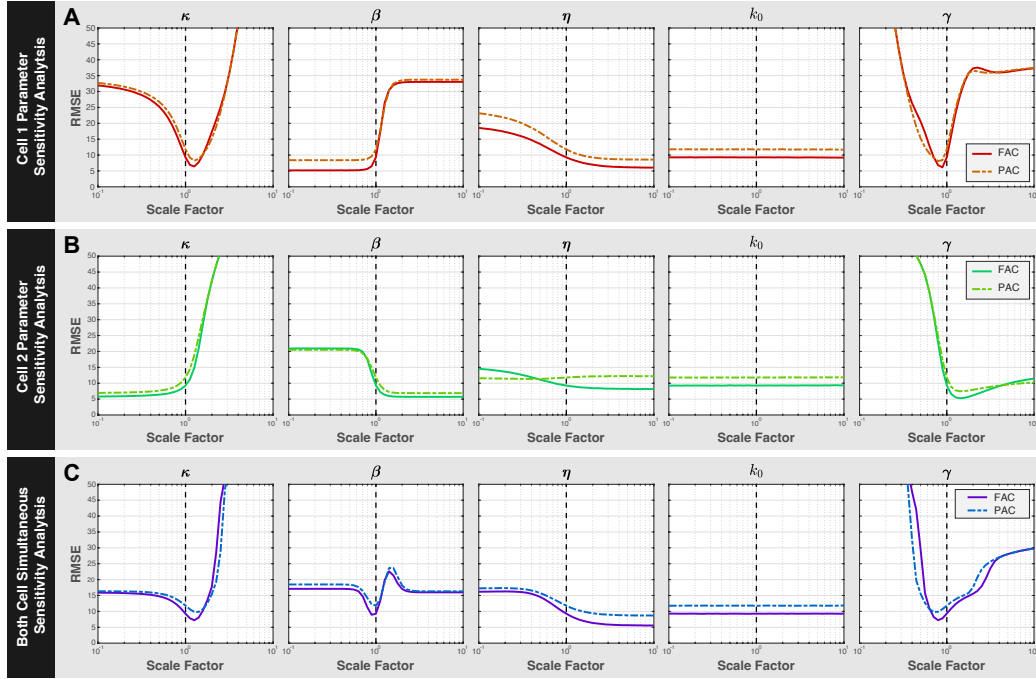$$r(t) = \mathbf{r}(\mathbf{x_n}(\mathbf{t})) \tag{20}$$

10

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**). Columns show each parameter in the model, rows show the cell which has its parameter perturbed.

3 Results

3.1 Effects of Parameter Errors or Extrinsic Uncertainties

Real stochastic processes often have unknown parameter values, and even genetically identical single-cells often exhibit heterogeneity in parameters due to external variations. Parameter estimates can be obtained, but are not perfect measurements due to measurement limitations of stochastic processes. To asses the sensitivity of control performance to parameter errors, sensitivity analysis was performed for each model parameter of each cell, then on both cells simultaneously. For each parameter perturbation, both the $\mathbf{u}^{\text{FAC}}$ and $\mathbf{u}^{\text{PAC}}$ controllers were examined across a parameter range from one-tenth to ten-fold of the original values. Figure 2A shows the results for these parameter sweeps when a single parameter in Cell 1 is fixed, and Fig. 2B shows the control performance as single parameters in Cell 2 are fixed while all other parameters are held constant. Finally, 2B shows the performance change when the corresponding parameters are changed in both Cell 1 and Cell 2 at the same time while keeping other parameters fixed.

Modification of parameters was found to produce a broad range of effects. 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.

Modification of parameters was found to produce a broad range of effects. Increasing β in Cell 1 was found to worsen performance, while increasing β in Cell 2 improved performance (Fig. 2A, B, second column). The effects of parameter modification were not always monotonic. For example, increasing κ in Cell 1 initially improved performance but reached a limit beyond which control performance degraded rapidly (Fig. 2A, left column). In contrast, perturbations in the parameter k_0 had insignificant effects on performance, even for relatively large perturbations (s=10) (Fig. 2A). Similarly, when parameters of both Cell 1 and Cell 2 were jointly changed, control performance

could be either improved or decreased (Fig. 2C).

Changing parameters had similar effects on control performance for both $\mathbf{u}^{\text{FAC}}$ (solid lines) and $\mathbf{u}^{\text{PAC}}$ (dashed lines). Parameters such as κ and γ reached minima that performed better than the original values, indicating potential for system optimization beyond the original design. However, altering k_0 had no noticeable impact, possibly due to its small original value requiring a larger scaling for observable changes.

In some parameters like κ or γ, the PAC controller showed greater robustness to parameter errors than the FAC controller, even outperforming it. However, this effect is due to the mismatch between assumed and actual parameters, and re-optimizing both controllers for the new parameters will always yield better control performance 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 an origional 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 [**?**]).

Stochastic fluctuations in a chemical process decrease with the inverse square root of the process scale [**?**]. As the system size increases, the dynamics converge towards a deterministic process, except for certain exceptional initial conditions lying on manifolds that separate attractors for different steady state behaviors [**?**]. This creates a tradeoff in noise-induced control, where noise enables differential control but is also deleterious to maintaining desired states once they have been achieved. The impact of reducing noise through increased system granularity on control performance remains unclear.

To shed light on this tradeoff, control performance scores for the $\mathbf{u}^{\mathrm{FAC}}$ and $\mathbf{u}^{\mathrm{PAC}}$ controllers were calculated at different levels of granularity (α) between 0.2 and 2.0. Figure 3 (A - F) 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 3G shows the trend of the performance 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 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
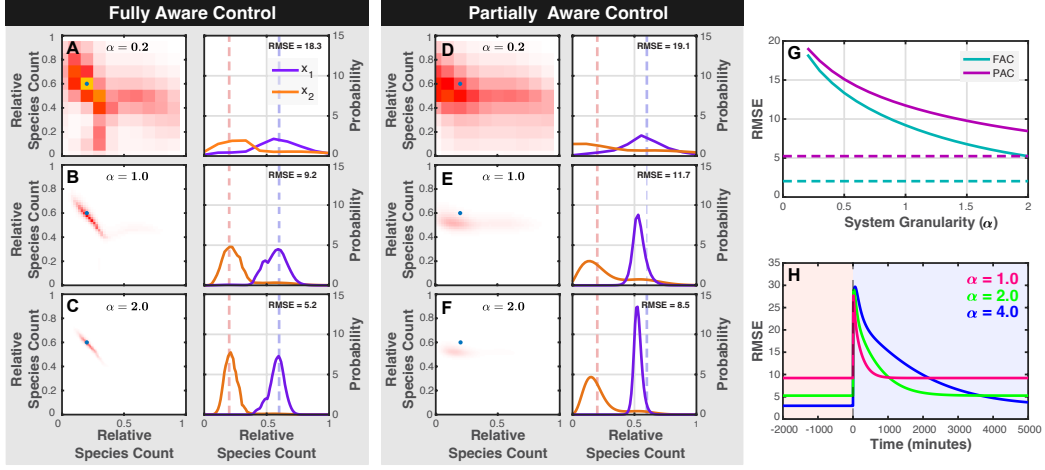
14

Figure 3: Systems with increased granularity are less noisy and have better control performance. (A-C) Joint (left) and marginal (right) distributions for the FAC shows increased control performance and tighter distributions as α increases from 0.2 (top row) to 2.0 (bottom row) . (D-F) Joint (left) and marginal (left) distributions using the PAC controller at different levels of granularity.

system scale, for all cases considered so far, we found that both controllers monotonically improved with increased α and that $\mathbf{u}^{\mathrm{FAC}}$ always outperforms the $\mathbf{u}^{\mathrm{PAC}}$.

To bypass this FSP limit, sixteen SSA simulations were simulated to steady state to sample the CME of a system with a significantly larger volume of $\alpha = 100$. Each SSA ran for 5×10^7 minutes, and the last 4×10^7 minutes were sampled for estimating the stationary distribution and calculating the performance score. The performance score estimates of this high granularity SSA, using the FAC and PAC controllers, were 4.13 and 27.5, respectively, shown as dashed lines in Fig. 3G. While it remains uncertain if further performance improvements can be achieved with larger system scales, these findings suggest that both controllers consistently improve with increased α, with $\mathbf{u}^{\mathrm{FAC}}$ consistently outperforming $\mathbf{u}^{\mathrm{PAC}}$ in all cases examined. From figure H, it can be seen that a smaller α creates faster response but higher baseline score due to intrinsic noise. When $\alpha = 4$, the system becomes much less noisy and has lower scores but takes longer time to respond.
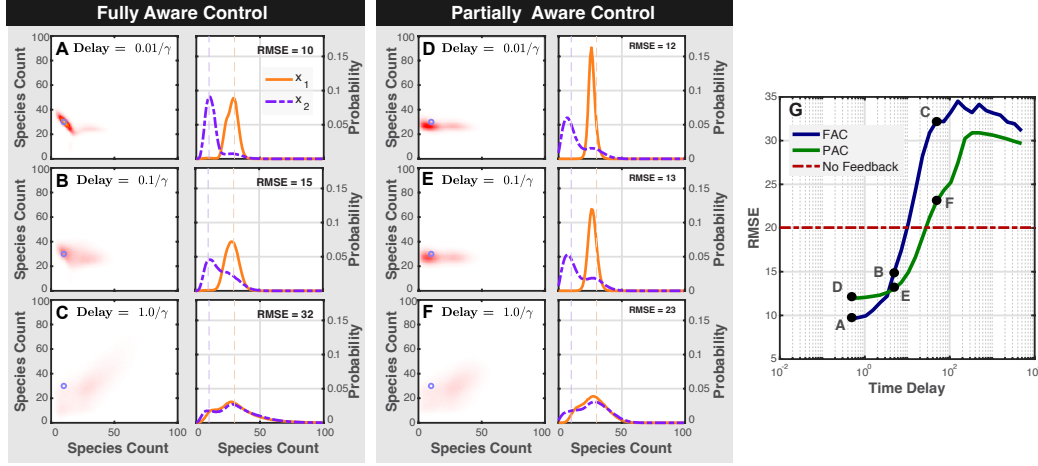
15

Figure 4: Stochastic simulations of the time delayed system show decreasing control performance using either the FAC or PAC. (A-C) Joint (left) and marginal distributions (right) of the at different levels of time delay for the FAC. (D-F) Joint (left) and marginal distributions(right) at three levels of time delay using the PAC. (G) plots of RMSE control performance versus time delay for both FAC and PAC.

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-enhanced controller $\mathbf{u}^{\text{FAC}}$ and $\mathbf{u}^{\text{PAC}}$, 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) the PAC outperforms the FAC (J =170 versus 219 at $\tau = 0.1/\gamma = 0.0203$ min).

This data show that time delays exceeding $0.07/\gamma = 0.0142$ minutes is detrimental to the FAC control performance while delays beyond $0.2/\gamma = 0.0406$ minutes are detrimental to the PAC controller. The choice of best controller depends on the level of time delay in the system. At extreme time delays, both controllers lose their asymmetry, resulting in significantly worse scores ($J = 1041$ and 931). The control performance of a no-feedback controller was measured which required no inputs Figure4(G) (red line) and achieved a score of $J = 402$. These findings indicate that, for large time delays, attempting to improve control with feedback is less effective than using a no-feedback approach.

By increasing α we can change the characteristic of a system from being more noise-like to more deterministic-like. Similarly, increasing τ can decrease the quality of the control input. Generally, we would expect that increasing α would improve control performance, while increasing τ would decrease control performance.

FAC steady state control performance was analyzed over a two dimensional domain of points (τ, α) using thirty-two stochastic simulations simulated to steady state to see if the positive effect granularity could undo the effect of negative effect of time delay. Instead, noise induced oscillations were surprisingly found at certain (τ, α) pairs and worse control performance was observed when system granularity increase for some τ. Joint probability distributions at certain (τ, α) pairs and their control performance can be seen in Fig 5(A-I). Control performance reduced as α increased when $\tau = 100$ but not at $\tau = 1$. FAC control performance Fig 5(J) over the domain of points shows a valley region at low τ, and a plateau region at large τ. This suggests that increasing system scale at large time delays is not beneficial to control performance. Trajectories at the edge of the valleyFig 5(J)(yellow star) showed oscillatory behavior Fig 5(K)and trajectories within the plateau show bursty behavior.

The control performance domain Fig 5(J) shows that low noise and high quality control inputs are both required to achieve good control. When noise is improved by increasing α the trajectories were observed going from very noisy to very oscillating which prevented improved control performance. De-
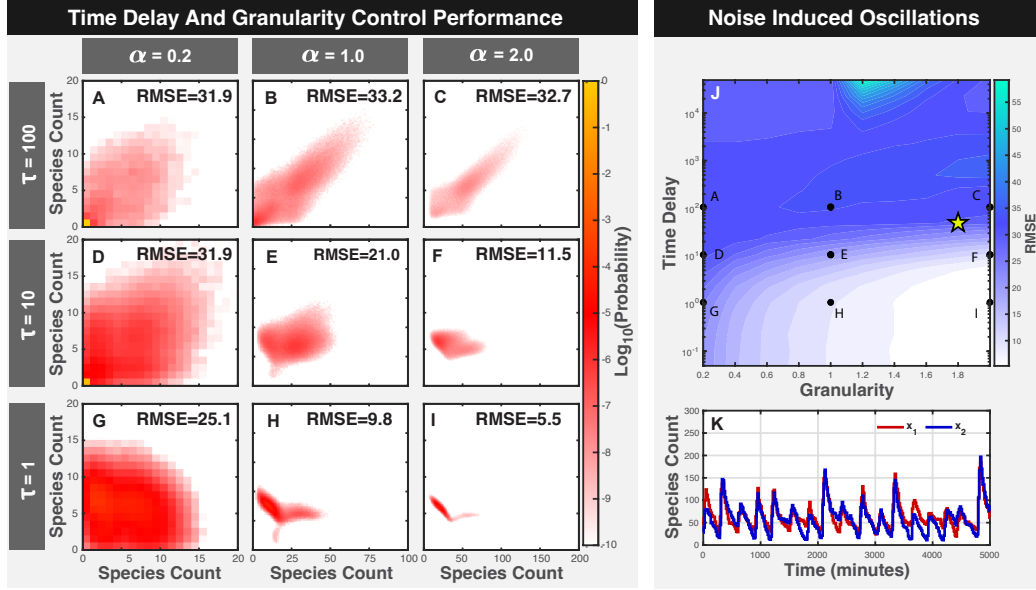
17

Figure 5: Stochastic simulations of the FAC driven system and PAC driven system show different failure modes at at different (τ, α) pairs. (A-I)Joint probability distributions at different $[\tau, \alpha]$ pairs. α across columns and τ across rows. (J) Control performance measured over a domain of (τ, α) pairs show good control performance when both α is large and τ small. (K) Trajectories where good control was expected but not observed show oscillations instead.
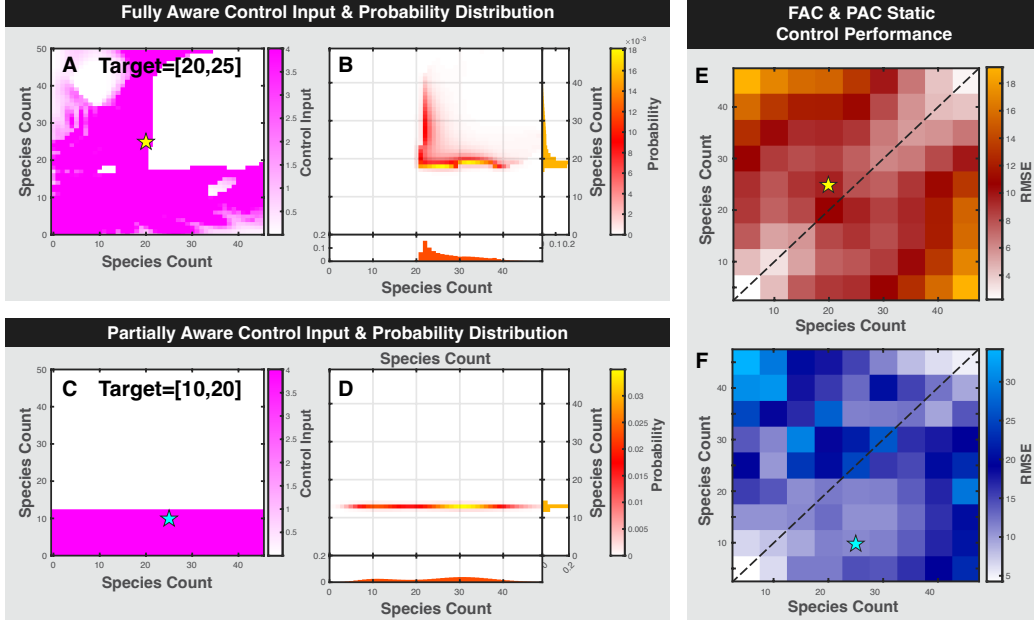
Figure 6: The FAC and PAC control methods were used to optimize control inputs and probability distributions for static set points in a 2D domain. Optimized control inputs and probability distributions were optimized using the FAC control method(A and B) or the PAC control method (C and D). (E) Domains of control performance for the FAC method over a region of static target points show regions of good control and symmetry along the dashed line. (F) Domain of control performance over a set of static control points for the PAC control method.

creases in α should make the system respond faster and be less susceptible to time delay, but also more noisy. These competing effects in control performance can be observed in Fig 5(J) which show that control performance is always poor at low α.

We analyzed the control performance over a domain of target points to find if some regions are easier to control than others. Stability analysis of the auto-regulated system using a moderate light input can create up to four stable points and one unstable point in the state space. These points give rise to basins which may attract probability more easily than other points. Control performance of the stochastic system was analyzed over a domain of different set points and were optimized using more observable (FAC) or less

observable (PAC) control.

Controllers were optimized over a two dimensional domain of target points $[T_1, T_2]$ between between five and forty-five using the FAC input method or the PAC input method. Figure 6 (A) shows the optimized control input of the system when the target is $T = [20, 25]$ after optimizing the FAC. Figure 6 (B) show the steady state probability distribution of the controlled system. Figure 6 (C and D) show the optimized controller and probability distributions using the partially aware control method. The unstable point in the rate equation matches the unstable point near $(20, 20)$ in this domain.

FAC and PAC control performance over the domain of static set points Fig6(E) and (F) showed regions of good control in the FAC between 10 and 30 but a smaller region in the PAC. The FAC control performance was found to be symmetric along the dashed black line. The cause was found to be that symmetric pairs had control inputs which were exact transposes of each other. The unstable point of the system is known to be near [20,20] which is also a point of poor control for the FAC system with respect to local control performance. Edge regions were difficult to control, likely due to the geometry of how control inputs move the probability.

We developed a method to control the dynamics of a system by optimizing steady state controllers for a predefined path $r(x)$ in a Piecewise manner and cycling through them at a fixed frequency. Thirty-two controllers were optimized along a pathway in order to target an in-sync reference point 7, a phase lagged reference point 7, or frequency separated reference point7(A). The synchronous signal is given by

$$r_1(x) = [20sin(2\pi x) + 10, 20sin(2\pi x) + 10] \tag{21}$$

the phase lagged signal is given by

$$r_2(x) = [20sin(2\pi x) + 10, 20cos(2\pi x) + 10] \tag{22}$$

and the frequency changed signal is given by

$$r_3(x) = [20sin(2\pi x) + 10, 20sin(6\pi x) + 10] \tag{23}$$

for $x \in [0, 1]$ using 32 steps. FSP simulations of cyclically driven systems were slowly driven at a frequency of $1e - 4$ seconds and their medians are shown in the Fig 7(B4, C4, and D4)(red line, blue line). The time series probability distributions of x_1 and x_2 can be seen as red clouds and blue
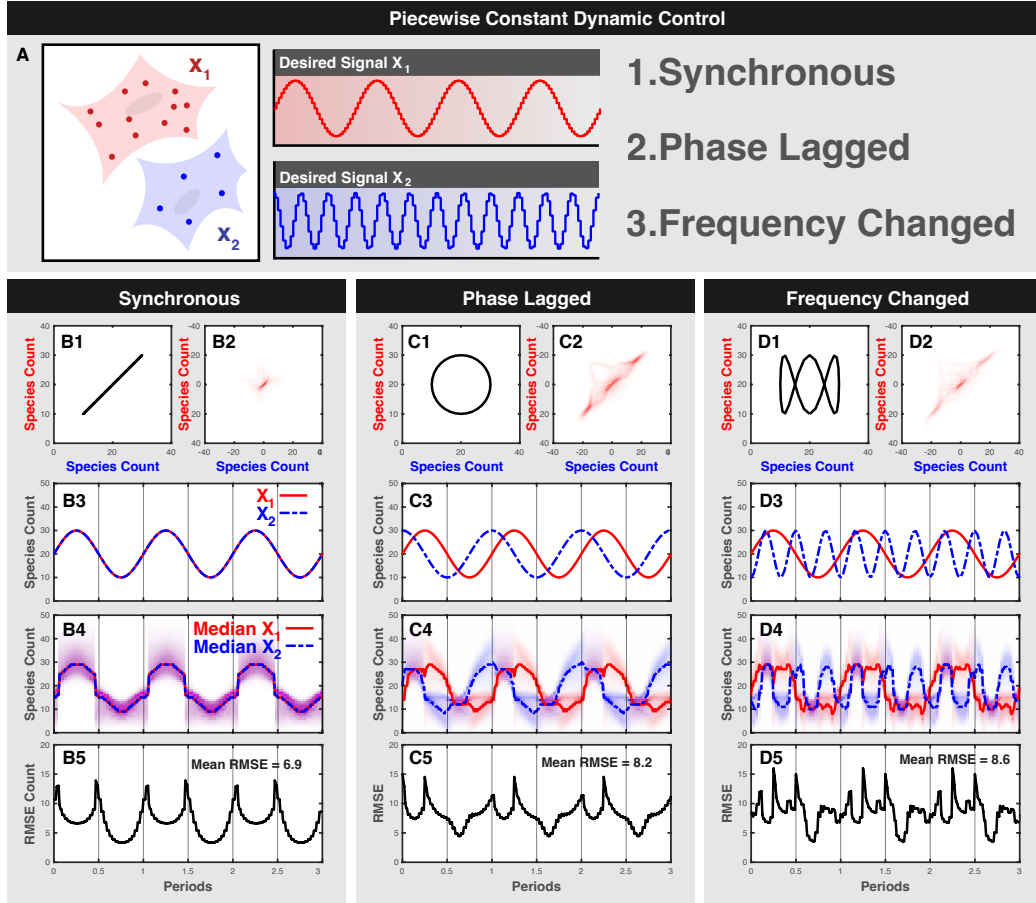
Figure 7: Finite state projections of the slowly driven system demonstrate good control performance under a variety of phase and frequency scenario. (A) Schematic of the control problem shows targeting the number of protein in two cells using a dynamically changing reference point by cycling through a set of individual stable points and under synchonous, phasse-lagged or frequency changes reference signal using only one control input. (B1-BN) shows the phase space of the reference phase space, error residuals, reference time series, FSP solution and RMSE respectively for the synchonous control method. (C1-C5) Shows the same data for the phase-lagged control input. (D1-D5) Shows the same analysis for the frequency-changed control.

21

clouds respectively. Regions with purple clouds are the exact overlap of red and blue. Error distributions of the system were taken by shifting the probability distribution from the reference point to the origin Fig 7(B2, C2, and D2). Root mean squared error of each simulation are shown in Fig 7(B5, C5,D5).

Suprisingly, the system is able to effectively control two signals to two different trajectories at different phase and frequency and using only a single control input despite the presence of noise Fig 7(D4). Simulations show that the synchronous control performed the best with an RSME of 6.9 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 8.2 and when driven at a different frequency the RMSE goes up to 8.6. Medians of the marginal distributions show good match with the desired reference signal in each control scheme Fig 7(B4,C4, and D4). TIme series trajectories of the RMSE show spikes in the control performance using any control strategy.

High-frequency driven systems often experience reduced control performance due to the system's slow response. However, the combination of both high-frequency driving and noise could be catastrophic for system control performance. Conversely, increasing granularity may improve control performance but alters the system's time scale.

To examine how driving frequency affects noise-exploiting control, Sixty-four SSA trajectories of the phase-separated dynamic controller were simulated over a two dimensional domain of points (α, f) for 16 cycles after reaching steady state. For each (α, f) pair the control performance of the steady state trajectories were calculated.

Fig 8(A) shows the desired reference of a system with $\alpha = 5.0$, and Fig 8(A) shows the mean of the system response when $f = 1e - 4$. Fig 8(C) Shows control performance over time as α is held at 5.0 and f increases from 1e-4 to 1e-2 to 1e-0 with an average RMSE of 6.4, 16.3, and 14.4 respectively. I creasing frequency while alpha is fixed at 0.5, 1.0 and 2.0 shows that control performance from 12 to 6.3.

Heat maps of control performance of (f, α) pairs showed the worst performance at moderate f and that control performance was best when both granularity was high and frequency was low.
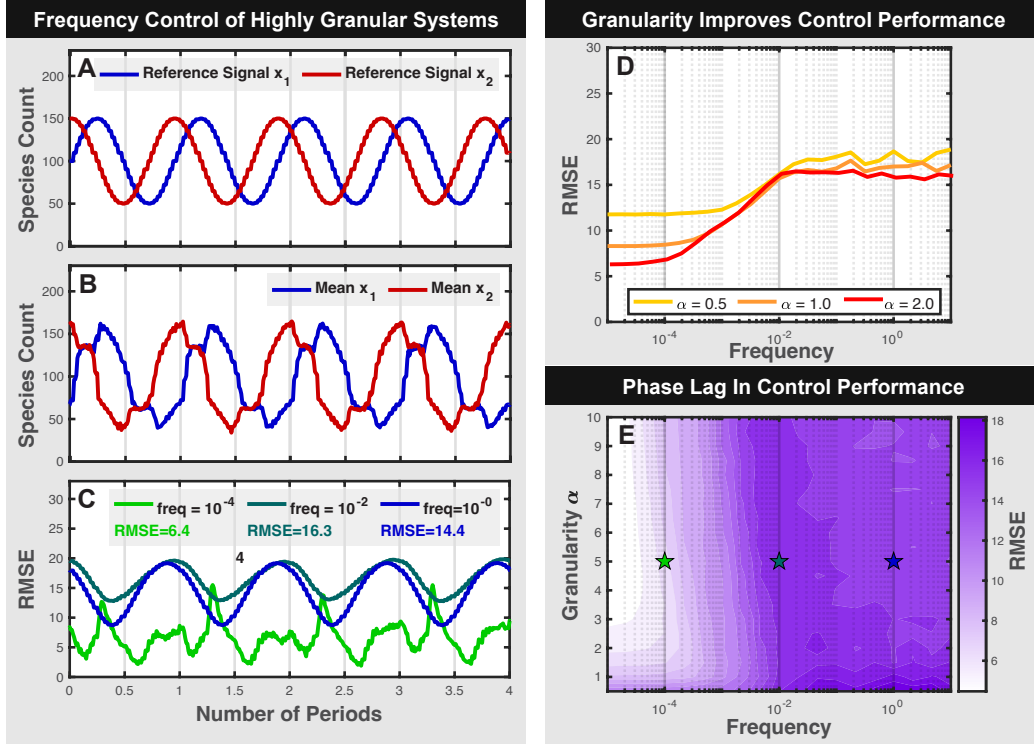
Figure 8: Control performance analyzed over a domain of f, α pairs show worst control performance at moderate frequencies near $1e-2$ Tracking reference signals when $\alpha = 5$ span a range between 50 and 150 species (A). Stochastic simulations driven using a phased lagged controller at $\alpha = 5$ and low frequency show tighter control compared to $\alpha = 1$ (B). Systems driven at moderate frequency show worse control performance than high frequency or low frequency (C). Control performance only of phase-lagged system only improves with increasing α and low f.

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.