Coupled Reaction Networks for Noise Suppression

Fangzhou Xiao[1], Meichen Fang[2], Jiawei Yan[3] and John C. Doyle[4]

***Abstract*— Noise is intrinsic to many important regulatory processes in living cells, and often forms obstacles to be overcome for reliable biological functions. However, due to stochastic birth and death events of all components in biomolecular systems, suppression of noise of one component by another is fundamentally hard and costly. Quantitatively, a widely-cited severe lower bound on noise suppression in biomolecular systems was established by Lestas *et. al.* in 2010, assuming that the plant and the controller have separate birth and death reactions. This makes the precision observed in several biological phenomena, e.g., cell fate decision making and cell cycle time ordering, seem impossible. We demonstrate that coupling, a mechanism widely observed in biology, could suppress noise lower than the bound of Lestas *et. al.* with moderate energy cost. Furthermore, we systematically investigate the coupling mechanism in all two-node reaction networks, showing that negative feedback suppresses noise better than incoherent feedforward achitectures, coupled systems have less noise than their decoupled version for a large class of networks, and coupling has its own fundamental limitations in noise suppression. Results in this work have implications for noise suppression in biological control and provide insight for a new efficient mechanism of noise suppression in biology.**

I. Introduction

Many important processes in living cells, such as gene expression, are intrinsically stochastic [1]. The effect of noise is further enhanced by several biological factors, such as the low copy number of many important molecules including DNA, and the fact that even controllers performing noise suppression consists of intrinsically stochastic molecular components themselves [1], [2], [3]. In particular, assuming that the controller and the plant dynamics are separate biochemical processes, Lestas *et. al.* [2] showed that there is a severe lower bound on the noise for the plant component of a chemical reaction network due to the intrinsic stochasticity of controller components in chemical reactions. Quantitatively, it states that the lower bound for noise is typically inversely proportional to the quartic root of the signaling rate, in contrast with a square root result that is familiar in electrical engineering and statistics. This bound implies that significant noise is inevitable, despite the regulatory mechanisms found in biology that could suppress noise [4]. Specifically, the bound implies that noise suppression in cells is usually prohibitively expensive, as reducing noise by 10 fold would require 10,000 fold increase in signaling rate, which corresponds to 10,000 fold faster production and degradation of molecules.

On the other hand, we observe almost deterministic precision in many biological processes, such as the synchronization of molecular oscillations in mammalian circadian clock [5], precise timing of gene expression dynamics in cell cycle [6] and the cell type differentiation through gene expression patterns during development [7]. This observation naturally urges one to seek biologically relevant low-cost noise suppression mechanisms that are beyond the scope of systems investigated in [2].

In this work we propose coupling as such a noise suppression mechanism that is omnipresent in biological processes. Coupling is used here to describe chemical reactions where the number of more than one chemical species are altered simultaneously, e.g., reactions converting one species into another one, or those producing one molecule each of two species simultaneously. Hence, coupling naturally goes beyond the assumption in [2] that the plant and the controller dynamics are separate biochemical processes, since the increase and/or decrease of a plant component and a controller component could be coupled in a reaction once we allow coupling.

Many natural and genetically engineered circuits in biology have a coupling interpretation. In bacteria, genes are commonly grouped into operons, so they are transcribed and regulated together [8]. Non-coding RNAs with regulatory functions such as microRNA and siRNA are commonly or designed to be in the same transcript with the mRNA they regulate [9], [10]. In metabolic networks, signaling cascades and allosteric regulation, one protein may switch from an inactivated state to an activated state by a reaction which may be regulated by the protein or its downstream products [8], [11]. Regulatory functions in synthetic biology are commonly implemented in binding proteins linked with functional domains, where the link is cut to transform the protein-domain complex into a protein and a domain separately, which could have different regulatory properties than the complex [12], [13].

The fact that specific coupling mechanisms, e.g. putting genes on the same operon, could influence the system's performance is commonly known. Coupling as a noise suppression mechanism is also rather intuitive from an information-theoretic point of view. Indeed, earlier studies [10], [14], [15], [16] have shown a few specific examples of coupled reactions could have less noise than the decoupled versions.

[1]Fangzhou Xiao is with Department of Biology and Biological Engineering, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91125, USA. xfz@caltech.edu

[2]Meichen Fang is with School of Life Sciences, Peking University, 5 Yiheyuan Rd., Beijing 100080, China. fangmeichen@pku.edu.cn

[3]Jiawei Yan is with Department of Systems Biology, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA. jiaweiyan@g.harvard.edu

[4]John Doyle is with Department of Control and Dynamical Systems, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91125, USA., doyle@caltech.edu

However, it is not known what are the general conditions for coupling to suppress noise, the fundamental limitations on noise suppression once we allow coupling (in the spirit of [2]), or how to use coupling when designing biochemical network architectures.

In this work we set off to provide some initial answers to these problems. After reviewing the notions of chemical reaction networks and methods to analyze them in Section II, we show how coupling could suppress noise through a simple example in Section III-A. We then demonstrate the power of coupling via a biologically plausible coupled system that suppresses noise below the lower bound of Lestas *et. al.* [2] in Section III-B. Following Lestas *et. al.* [2], we establish a similar bound for coupled reaction networks in Section III-C to better understand how it compares with decoupled systems and when is coupling most useful. Lastly, in Section III-D, we investigate all possible two-node chemical reaction network architectures to gain insight on the advantage of coupled networks versus decoupled ones, feedback versus feedforward architectures, limitations of coupling, and how to choose network architectures to maximize coupling's effect on noise suppression.

II. Background

A. Chemical Reaction Networks

Here we provide a minimum background on stochastic chemical reaction networks (CRNs) and relevant analysis methods. For more details, see [3], [17], [18].

A chemical reaction network (CRN) is a collection of reactions of the following form:

$$\boldsymbol{x} \xrightarrow{r_k(\boldsymbol{x})} \boldsymbol{x} + \boldsymbol{d}_k,$$

where $\boldsymbol{x} \in \mathbb{Z}^n_{\geq 0}$ is the state vector describing the number of molecules of the n chemical species in the system, $k = 1, ..., m$ indexes m reactions, and $r_k : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is the propensity function which describes the rate of reaction k. Technically, the system dynamics is described as a continuous-time Markov jump process with $\boldsymbol{x} \in \mathbb{Z}^n_{\geq 0}$ as states of the system, and $r_k(\boldsymbol{x})\Delta t$ describes the probability that, if the system is in state $\boldsymbol{x}$ at time t, it will jump to state $\boldsymbol{x} + \boldsymbol{d}_k$ within time Δt, for small Δt. Note that even though the mass action law, where r_k are always a special form of polynomial of $\boldsymbol{x}$, is commonly assumed [19], we do not make that assumption here.

The stochastic description of the system dynamics is described in terms of the chemical master equation (CME), which describes the evolution of the probability distribution for molecular counts $\boldsymbol{x}$ [20]:

$$\frac{d}{dt}P(\boldsymbol{x}, t) = \sum_k \left(r_k(\boldsymbol{x} - \boldsymbol{d}_k)P(\boldsymbol{x} - \boldsymbol{d}_k, t) - r_k(\boldsymbol{x})P(\boldsymbol{x}, t) \right)$$

where $P(\boldsymbol{x}, t)$ is the probability that the system has $\boldsymbol{x}$ number of molecules at time t.

For the mean, we have

$$\begin{aligned}\frac{d}{dt}\langle x_i \rangle =& \frac{d}{dt}\sum_{\boldsymbol{x}} x_i P(\boldsymbol{x}, t) = \left\langle \sum_k d_{ik} r_k(\boldsymbol{x}) \right\rangle \\ =& \left\langle \sum_{k:d_{ik}>0} d_{ik} r_k(\boldsymbol{x}) \right\rangle - \left\langle \sum_{k:d_{ik}<0} |d_{ik}|\, r_k(\boldsymbol{x}) \right\rangle \\ \equiv& \left\langle R_i^+(\boldsymbol{x}) \right\rangle - \left\langle R_i^-(\boldsymbol{x}) \right\rangle,\end{aligned}$$

where the last equality was used as the definition for R_i^+ and R_i^-. R_i^+, the birth flux of x_i, is the summed rate of all reactions that increases x_i, while R_i^-, the death flux, is the summed rate of all reactions that decreases x_i. Note that, at steady state, we have $\langle R_i^+ \rangle = \langle R_i^- \rangle$.

It is worth noting that, in the limit where the reaction volume is large while the number of molecules per volume remains finite (made precise in [21]), the system dynamics can be described by a deterministic rate equation:

$$\frac{d\tilde{x}_i}{dt} = \sum_{k=1}^{m} d_{ik} r_k(\tilde{\boldsymbol{x}}) = R_i^+(\tilde{\boldsymbol{x}}) - R_i^-(\tilde{\boldsymbol{x}}),$$

where $\tilde{\boldsymbol{x}} \in \mathbb{R}^n_{\geq 0}$ is the continuous concentrations of chemical species, instead of discrete molecular counts $\boldsymbol{x} \in \mathbb{Z}^n_{\geq 0}$.

Through similar calculations as in the equation for the mean, we see that, at steady state, we have the following equation for covariance:

$$\mathrm{Cov}\left(x_i, R_j^- - R_j^+\right) + \mathrm{Cov}\left(x_j, R_i^- - R_i^+\right) = \sum_k d_{ik} d_{jk} \langle r_k \rangle.$$

One major achievement of the theoretical investigations in [3] is the re-writing of the above equation using physically interpretable quantities. The re-written equation is as follows:

$$\begin{aligned}&\frac{1}{\tau_j}\frac{\mathrm{Cov}\left(x_i, R_j^- - R_j^+\right)}{\langle x_i \rangle \langle R_j \rangle} + \frac{1}{\tau_i}\frac{\mathrm{Cov}\left(x_j, R_i^- - R_i^+\right)}{\langle x_j \rangle \langle R_i \rangle} \\ &= \frac{1}{\tau_i}\frac{\langle s_{j|i} \rangle}{\langle x_j \rangle} + \frac{1}{\tau_j}\frac{\langle s_{i|j} \rangle}{\langle x_i \rangle} := D_{ij} = \frac{\sum_k d_{ik} d_{jk} \langle r_k \rangle}{\langle x_i \rangle \langle x_j \rangle}.\end{aligned} \tag{1}$$

Here τ_i, the average life time of an x_i molecule, and $\langle s_{i|j} \rangle$, the average step sizes or co-step sizes, are introduced. As these are key concepts utilized for this work, they are explained in detail below.

Step sizes. For $i = j$, we define the average step size of x_i as $\langle s_{i|i} \rangle \equiv \sum_k |d_{ik}|\, \rho_{ik}$, where $\rho_{ik} = \frac{|d_{ik}|\langle r_k \rangle}{\sum_{k'} |d_{ik'}| \langle r_{k'} \rangle}$, the probability that when x_i changes, that change comes from reaction k. The notation $\langle s_{i|i} \rangle$ signifies that this is average step size of x_i conditioning on that x_i changes. For example, if x_1 has only one birth reaction $x_1 \to x_1 + 1$ and one death reaction $x_1 \to x_1 - 10$, then $\langle s_{1|1} \rangle = \frac{1+10}{2}$, because birth and death fluxes are always equal at steady state.

For $i \neq j$, the co-step size is defined as $\langle s_{j|i} \rangle \equiv \sum_k \rho_{ik}\, |d_{jk}|\, \mathrm{sgn}\{d_{ik} d_{jk}\}$. So $\langle s_{j|i} \rangle$ is the average change of species x_j conditioning on that x_i changes simultaneously in these reactions. Note that this could be positive or negative. From the definition, we see that the co-step size $\langle s_{i|j} \rangle$ captures whether birth and death of x_i and x_j are coupled in some reaction. For example, if the only reaction that have simultaneous changes to x_1 and x_2 is $(x_1, x_2) \to (x_1 - 1, x_2 + 1)$, i.e. one x_1 becomes one x_2, and x_2 have no other birth reactions, then $\langle s_{1|2} \rangle = -\frac{1}{2}$. It is negative because when x_1 decreases, x_2 increases. It is divided by 2 because this reaction accounts for all of x_2's birth, therefore half of x_2's changes.

Note that the co-step sizes and the birth and death fluxes

are related to each other: $\langle s_{j|i}\rangle \langle R_i\rangle = \langle s_{i|j}\rangle \langle R_j\rangle$.

Life times. For general stationary stochastic processes describing the increase and decrease of a quantity, e.g. molecule count, we can define the average life time as the average time that one molecule lasts before its degradation. Little's law in queueing theory then states that the average life time is equal to the ratio between the average number of such molecules and its average birth or death rate, regardless of how nonlinear the reaction rates are. This could be intuitively understood through the following non-rigorous calculation, where $\int_0^T x_i(t)dt$ is the sum of life time of all x_i molecules in T, and $\int_0^T R_i^+(t)dt$ is the total number of x_i molecules born in T:

$$\tau_i \equiv \frac{\lim_{T\to\infty}\int_0^T x_i(t)dt}{\lim_{T\to\infty}\int_0^T R_i^+(t)dt} = \frac{\lim_{T\to\infty}\langle x\rangle T}{\lim_{T\to\infty}\langle R_i^+\rangle T} = \frac{\langle x_i\rangle}{\langle R_i^\pm\rangle}.$$

Therefore, we have $\tau_i = \frac{\langle x_i\rangle}{\langle R_i^\pm\rangle}$, the average life time or the time scale of x_i.

B. Linear Noise Approximation

We see that the covariance equation (1) involves terms of the form $\mathrm{Cov}\left(x_i, R_j^\pm\right)$, which would involve higher order moments if $R_j^\pm$ is nonlinear. The resulting system of moment equations is generally infinite dimensional and hard to solve. This naturally calls for moment closure methods to approximately solve this system [22]. One particularly simple and analytically tractable method is linear noise approximation (LNA). LNA could be derived by assuming that the mean $\langle \boldsymbol{x}\rangle$ is close to a fixed point so that $\langle \boldsymbol{x}\rangle = R_j^+(\langle \boldsymbol{x}\rangle)$, and by simply approximating birth and death rates by their first order Taylor expansion: $R_j^-(\boldsymbol{x}) \approx R_j^-(\langle \boldsymbol{x}\rangle) + \sum_{\ell=1}^m \frac{\partial R_j^-(\langle \boldsymbol{x}\rangle)}{\partial x_\ell}\Delta x_\ell$. Define the logarithmic gains $H_{jl} \equiv \left(\frac{\partial R_j^-(\langle \boldsymbol{x}\rangle)}{\partial x_\ell}\frac{\langle x_\ell\rangle}{\langle R_j^-\rangle} - \frac{\partial R_j^+(\langle \boldsymbol{x}\rangle)}{\partial x_\ell}\frac{\langle x_\ell\rangle}{\langle R_j^+\rangle}\right)$ [17], equation (1) becomes

$$\frac{1}{\tau_j}\sum_{\ell=1}^m H_{jl}\eta_{il} + \frac{1}{\tau_i}\sum_{\ell=1}^m H_{il}\eta_{jl} = D_{ij}.$$

In matrix form, we have

$$\boldsymbol{M}\boldsymbol{\eta} + \boldsymbol{\eta}\boldsymbol{M}^\intercal = \boldsymbol{D}, \tag{2}$$

where $\eta_{ij} = \eta_{ji} = \frac{\mathrm{Cov}(x_i,x_j)}{\langle x_i\rangle\langle x_j\rangle}$, $D_{ij} = D_{ji} = \frac{1}{\tau_i}\frac{\langle s_{j|i}\rangle}{\langle x_j\rangle} + \frac{1}{\tau_j}\frac{\langle s_{i|j}\rangle}{\langle x_i\rangle} = \frac{\sum_k d_{ik}d_{jk}\langle r_k\rangle}{\langle x_i\rangle\langle x_j\rangle}$, $M_{ij} = \frac{H_{ij}}{\tau_i}$. Note that an internal relation needs to be utilized when solving this system of equations:

$$\frac{\langle s_{i|j}\rangle}{\langle x_i\rangle\tau_j} = \frac{\langle s_{j|i}\rangle}{\langle x_j\rangle\tau_i}, \tag{3}$$

which is derived from the relation $\langle s_{i|j}\rangle\langle R_j\rangle = \langle s_{j|i}\rangle\langle R_i\rangle$.

Equation (2) is called the fluctuation dissipation theorem in the statistical physics community [23] and Lyapunov equation in the control community [24]. Note that the normalized covariance $\boldsymbol{\eta}$ is a solution to this Lyapunov equation shows that it is finite if and only if the linearized deterministic system $\frac{d}{dt}\boldsymbol{x} = \nabla_{\langle x\rangle}(\boldsymbol{R}^+ - \boldsymbol{R}^-)\boldsymbol{x}$ is asymptotically stable, i.e. $-\boldsymbol{M}$ is Hurwitz. Hence there exists a general relationship between deterministic stability and stochastic variance [25].

As LNA is the main tool of analysis in this work, more comments on the effectiveness of LNA is in order. Although LNA was historically derived as a second order system size approximation of the chemical master equation, resulting in a Fokker-Planck equation with a stochastic differential equation interpretation and a Gaussian steady state distribution [26], this is not necessary. LNA, as well as higher order approximations, could be derived as Taylor expansions on the moment equations [27]. In fact, it could be shown that LNA could be interpreted as a linear propensity CRN approximation that preserves discreteness and non-negativity of the state variables (see Section V-B). So violating discreteness and non-negativity are not valid grounds for rejecting the usefulness of LNA. LNA fails just like how deterministic linearization fails: when nonlinearity significantly influences the state trajectory.

III. Results

A. Coupling could reduce noise

Here we illustrate how coupling could reduce noise through one simple example with very explicit analysis.

Consider the following feedforward network, called a "sniffer" system in biological literature [28].

$$\begin{aligned} x_1 &\xrightarrow{w} x_1+1, \quad & x_2 &\xrightarrow{w} x_2+1,\\ x_1 &\xrightarrow{kx_1x_2} x_1-1, \quad & x_2 &\xrightarrow{x_2/\tau_2} x_2-1. \end{aligned} \tag{4}$$

It consists of two components x_1 and x_2, both activated by an external input w, while x_2 acts as an enzyme that catalyze the degradation of x_1. The system (4) does not couple the birth events of x_1 and x_2, so they are produced with rate w through separate reactions. The corresponding coupled version of system (4) keeps the last two death reactions the same while substituting the following reaction for the first two birth reactions:

$$\{x_1, x_2\} \xrightarrow{w} \{x_1+1, x_2+1\}. \tag{5}$$

For both coupled and decoupled versions, we see that the birth rates of x_1 and x_2 are $R_1^+ = R_2^+ = w$, their death rates are $R_1^- = kx_1x_2$, $R_2^- = x_2/\tau_2$, and their step sizes are $\langle s_{11}\rangle = \langle s_{22}\rangle = 1$. For the coupled version, the co-step sizes are $\langle s_{12}\rangle = \langle s_{21}\rangle = \frac{1}{2}$, indicating half of the flux are through a coupled reaction where x_1 and x_2 are increased or decreased with the same number, while both co-stepsizes are 0 for the decoupled version. Applying LNA (see Section II), the steady states are $\langle x_1\rangle = \tau_2/k, \langle x_2\rangle = w/\tau_2$ and the time scales are $\tau_1 = 1/(k\langle x_2\rangle) = \frac{\tau_2}{kw}$. The gain matrix is $\boldsymbol{H} = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$. The diffusion matrix $\boldsymbol{D}$ in equation (2) are the following:

$$\boldsymbol{D}_c = \begin{bmatrix}\frac{2}{\tau_1\langle x_1\rangle} & \frac{1}{\tau_2\langle x_1\rangle}\\ \frac{1}{\tau_1\langle x_2\rangle} & \frac{2}{\tau_2\langle x_2\rangle}\end{bmatrix}, \quad \boldsymbol{D}_d = \begin{bmatrix}\frac{2}{\tau_1\langle x_1\rangle} & 0\\ 0 & \frac{2}{\tau_2\langle x_2\rangle}\end{bmatrix},$$

where $\boldsymbol{D}_c$ is for the coupled case and $\boldsymbol{D}_d$ is for the decoupled case.

Solving equation (2) with $\boldsymbol{D} = \boldsymbol{D}_d$ for the decoupled

case, we have

$$\eta_{22} = \frac{1}{\langle x_2 \rangle}, \qquad \eta_{12,d} = -\frac{\tau_2}{\tau_1 + \tau_2}\frac{1}{\langle x_2 \rangle},$$
$$\eta_{11,d} = \frac{1}{\langle x_1 \rangle} - \eta_{12,d} = \frac{1}{\langle x_1 \rangle} + \frac{\tau_2}{\tau_1 + \tau_2}\frac{1}{\langle x_2 \rangle}.$$

We see that the noise of x_1 in the decoupled case, $\eta_{11,d}$, can be decomposed into two parts: the first term $\frac{1}{\langle x_1 \rangle}$ that is the intrinsic noise of x_1 doing birth-death by itself, and the second term that is the noise of x_2 carried over to x_1 through its regulation on x_1.

In comparison, solving equation (2) with $\boldsymbol{D} = \boldsymbol{D}_c$ for the coupled case, with the additional constraint $\tau_2 \langle x_1 \rangle = \tau_1 \langle x_2 \rangle$ which is a consequence of equation (3), we have the following:

$$\eta_{22} = \frac{1}{\langle x_2 \rangle}, \qquad \eta_{12,c} = 0, \qquad \eta_{11,c} = \frac{1}{\langle x_1 \rangle}.$$

Note that $\eta_{11,c}$ is the same as that of a simple birth-death process of x_1 by itself, with no interaction with x_2 at all. The noise for the decoupled case is larger than the coupled case by $\eta_{11,d} - \eta_{11,c} = \frac{\tau_2}{\tau_1+\tau_2}\eta_{22}$, x_2's noise passed onto x_1 through their interaction. This indicates that the noise contributed to x_1 from the interaction with x_2 in the decoupled case is cancelled out here due to coupling.

This example shows that coupling could indeed reduce noise. In fact, similar circuits have been analyzed in [14], [10] that report specific RNA circuit designs with coupling could suppress gene expression noise.

B. Coupling suppresses noise below Lestas et. al. bound

We show here through a biologically plausible example that coupled systems could have noise below the bound described in [2].

Consider the following system:

$$\begin{gathered} x_1 \xrightarrow{x_1/\tau_1} x_1 - 1, \qquad x_2 \xrightarrow{k/x_1} x_2 + 1, \\ \{x_1, x_2\} \xrightarrow{x_2/(n\tau_2)} \{x_1 + 1, x_2 - n\}. \end{gathered} \tag{6}$$

Here n molecules of x_2 can be converted to 1 molecule of x_1, resulting in a negative coupled reaction. Besides this, x_1 has first order death reaction by itself, while x_2 is produced through a reaction that is suppressed by x_1.

When $n = 1$, the coupled reaction could be an enzyme switching from an inactive state x_2 into an active state x_1, which in turn suppresses the production of this enzyme x_2. Alternatively, it could be a binding protein in complex with a functional domain x_2 being digested into separate parts, where the free binding protein x_1 suppresses the production of this complex x_2. When $n > 1$, the coupled reaction could be protein subunits x_2 forming a protein complex x_1 involving n such subunits (e.g. $n = 2$ corresponds to a dimer), while x_1 has one of its active functions the suppression of subunit x_2's production. Therefore, this system is biologically plausible with potential straight-forward implementations through tools in synthetic biology.

To compare with the Lestas *et. al.* bound, we need to delve into the details of their work [2]. Lestas *et. al.* considered reaction networks of the following form:

$$\begin{aligned} x_1 &\xrightarrow{u[\mathcal{I}_t(x_2)]} x_1 + 1, \qquad & x_1 &\xrightarrow{\gamma x_1} x_1 - 1, \\ x_2 &\xrightarrow{f(x_1)} x_2 + 1, \qquad & x_2 &\xrightarrow{g(x_2)} x_2 - 1, \end{aligned} \tag{7}$$

where x_1 is the plant species whose noise is to be controlled, while x_2 is the signaling controller species whose information on x_1 is the only source of information on x_1 we could use to suppress its noise. The signaling of x_2 is through its birth events, as the birth rate f is dependent on x_1. For example, if $f(x_1) = x_1$, so x_1 catalyzes the production of x_2, then we would estimate x_1 to be large if we observe a high density of x_2's birth events. Therefore, information about x_1's abundance could be extracted through a trajectory of x_2's birth events. Because x_2's death rate g does not depend on x_1, the death events does not give any information about x_1. So we ignore death events and focus on x_2's birth events. On the other hand, with the information about x_1 extracted through observations on x_2, we try to suppress x_1's noise by controlling its (possibly non-Markovian) production rate, $u[\mathcal{I}_t(x_2)]$, which could be an arbitrary non-anticipatory functional allowing dependence of all the information $\mathcal{I}_t(x_2)$ of past x_2. In control theory terms, x_1 is the plant, birth events of x_2 with signaling rate f is the sensor, and x_1's birth rate u is the controller actuation.

Lestas *et. al.* then proceeded to use sensor and actuator separation in order to first bound the channel capacity in terms of f and then bound the noise in terms of channel capacity. The bounds are the following:

$$C = \langle f \rangle \log\left\{1 + \frac{\mathrm{Var}\{f\}}{\langle f \rangle^2}\right\}, \quad \eta_{11} \geq \frac{1}{\langle x_1 \rangle}\frac{1}{C\tau_1 + 1}, \tag{8}$$

where C is an achievable upper bound for the channel capacity of x_2's birth events assuming finite mean and variance of f, and τ_1 is the average life time of x_1 as defined in Section II, which is equal to $1/\gamma$ for system (7). If we can write the mean and variance of f in terms of those of x_1 (e.g. when $f \propto x_1$), then these two bounds could be combined to produce a lower bound of x_1 in terms of only x_1's mean, variance, and time scales. Importantly, when f is a linear function of x_1, i.e., $f = \alpha x_1$, this bound is:

$$\eta_{11} \geq \frac{1}{\langle x_1 \rangle}\frac{2}{1 + \sqrt{1 + N_2/N_1}}, \tag{9}$$

where $N_1 = \langle x_1 \rangle$ and $N_2 = \alpha \langle x_1 \rangle \tau_1$ are the effective signaling rates, defined as the number of birth events of x_1 and x_2 during an average life time of x_1, i.e., τ_1, respectively. This is the quartic bound for the coefficient of variation ($\sqrt{\eta_{11}}$) mentioned in Section I.

Going back to our example (6), we see that the number of x_2 molecules consumed in the coupled reaction, n, is equal to the relative signaling rate $\frac{N_2}{N_1}$ as defined above. So we expect to see better noise suppression with increasing n. The birth rate satisfies $f = k/x_1$, which is not a linear function of x_1, so we need to use equation (8) instead of the linear bound (9). In simulation, we need to estimate mean and variance of f to calculate C, which in turn enables calculation for lower bound of η_{11} in equation (8). It is important to note that we violated none of the assumptions about (7) in Lestas *et. al.*

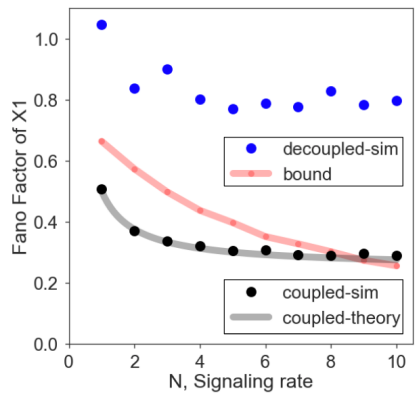

Fig. 1. Noise of example (6) in Section III-B compared to Lestas *et. al.* bound in equation (8) [2]. The y axis is the variance divided by mean of x_1, or its Fano factor. The x axis is the signaling rate $N = N_2/N_1$, taking integer values from 1 to 10. The Gillespie simulation results of system (6) are the black dots (coupled-sim), in good agreement with the theoretical results based on LNA shown as the light grey curve (coupled-theory). The theoretical bound of equation (8) with estimated capacity is shown as the light red curve (bound). The same simulation results for the decoupled version of system (10) are the blue dots (decoupled-sim). Parameters used for simulation: $k = 10000$,$\tau_1 = 100$,$\tau_2 = 1$.

bound (8) except allowing coupling among reactions.

Using LNA, we could analytically solve system (6) to have $\eta_{11} = \frac{1}{\langle x_1 \rangle}\frac{n+1}{4n}$ in the fast controller dynamics limit, where $\tau_2 \ll \tau_1$ (see Section III-D). The simulation result comparing the simulated noise of x_1, its LNA analysis, and the Lestas *et. al.* bound in equation (8) is shown in Figure 1. We see that the noise is indeed well below the theoretical bound of Lestas *et. al.* for not too large n, which is the more biologically plausible parameter regime.

We also did simulation for the decoupled version of system (6), where the following two reactions substituted the coupled reaction:

$$x_1 \xrightarrow{x_2(n\tau_2)} x_1 + 1, \qquad x_2 \xrightarrow{x_2(n\tau_2)} x_2 - n. \tag{10}$$

The simulation results for this decoupled version are the blue dots in Figure 1, showing that it is indeed above the Lestas *et. al.* bound, as the decoupled case satisfied all of the assumptions in their work [2].

It should also be noted that although this specific example (6) could achieve noise below the bound of [2], the scaling of noise in N in this system is actually worse than quartic root because it has a lower bound of $\frac{1}{4}$, as is shown by the black curve's convergence to $\frac{1}{4}$ in Figure 1 as well as the LNA calculation. We will see in Section III-D that there exists in general a non-zero lower bound on the part of noise that can be suppressed through coupling. In other words, once the architecture of a system is determined, there is a lower bound on how much noise can be reduced by coupling. For a direct comparison, in the following section, we establish a lower bound on noise suppression when coupling reactions are allowed using information theoretic methods in [2].

C. Bound on noise suppression with coupling

To better understand how coupling changes the fundamental limit on noise suppression, here we establish a bound similar to (9) that allows for coupled reactions. For detailed derivation, see V-C.

Let us first understand how coupled reactions go beyond bound (9). Bound (9) assumes that the birth events of x_2 is a Poisson channel whose rate is x_2's birth rates. However, when the birth or death events of x_1 and x_2 are coupled, the mutual information between the two increases, since a change of x_2 due to this coupled reaction always comes with a known change of x_1 at exactly the same time. Hence the key to establish a bound that includes coupling is in bounding the mutual information through this coupled birth-death.

Consider a coupled reaction in place of the decoupled $x_2 \to x_2 + 1$ in equation (7):

$$\{x_1,\ x_2\} \xrightarrow{w(x_1)} \{x_1 - 1,\ x_2 + 1\}, \tag{11}$$

which could be regarded a simple conversion reaction where one x_1 is converted into one x_2. We could split this channel into two Poisson channels, where one is $x_2 \to x_2 + 1$ with rate $w(x_1)$ just like the decoupled case, and one is $x_1 \to x_1 - 1$ with rate that is a function of x_1 and x_2, such that it becomes infinitely fast whenever it sees x_2 change so as to bring x_1 to match the change of x_2, and once x_1 changed it becomes 0. In other words, the $x_1 \to x_1 - 1$ reaction is to track the change of x_2 instantaneously.

With this consideration, we could bound the channel capacity of reaction (11) by $\langle w \rangle \ln(1+1/\langle w \rangle)+\mathrm{Var}\{w\}/\langle w \rangle$. The first term comes from using equation (8) to bound the channel capacity of the $x_2 \to x_2+1$ reaction with rate $w(x_1)$. The second term comes from the infinitely fast reaction $x_1 \to x_1 - 1$ mimicing the coupling. If we assume first order reactions $w(x_1) = \beta x_1$, then we can derive a bound in a form similar to Lestas *et. al.* bound in equation (9):

$$\eta_{11} \geq \frac{1}{\langle x_1 \rangle}\frac{2}{1 + Q + \sqrt{(1+Q)^2 + 4N_2/N_1}} \tag{12}$$

where $Q = N_2 \ln(1+\tau_1/N_2) \leq \tau_1$, and $N_2 = \langle w(x_1) \rangle \tau_1 = \beta \langle x_1 \rangle \tau_1$ now is the signaling rate of the coupled reaction. Note that $\frac{N_2}{N_1} = \beta\tau_1$ here. Compared to the decoupled bound (9), equation (12) contains one more term $Q \geq 0$ due to coupling. Note that this bound depends on the absolute value of N_2 in addition to the relative signaling rate N_2/N_1.

This means the effect of noise suppression depends on both $N_1 = \langle x_1 \rangle$ and τ_1. To unravel this dependence, we first notice that if τ_1 is small, then $Q \approx \tau_1$ is small as well, so the bound (12) approaches the decoupled bound (9). Furthermore, since $\frac{N_2}{N_1} \propto \tau_1$, this means signaling rate is small in general, yielding $\eta_{11} \geq \frac{1}{\langle x_1 \rangle}$, the Poisson noise lower bound. In other words, a small τ_1 indicates that noise suppression below Poisson is infeasible. This is also intuitively clear, as small τ_1 means x_1 or plant dynamics is so fast that our controller is simply too slow to be effective.

If τ_1 is not too small, then Q is of order N_2. So whether coupling effect is significant depends on how N_2 compares with $\frac{N_2}{N_1}$. If $N_1 = \langle x_1 \rangle$ is small, then relative signaling rate $\frac{N_2}{N_1}$ dominates while coupling has little influence, and the bound (12) approaches the decoupled bound (9) again, with minimum CV (i.e., $\sqrt{\eta_{11}}$) bounded by quartic root of $\frac{N_2}{N_1}$. If

$N_1 = \langle x_1 \rangle$ is not too small, then N_2 is comparable with $\frac{N_2}{N_1}$ and coupling is advantageous. Indeed, in the other extreme where $\frac{N_2}{N_1} \ll N_2$, we have $\eta_{11} \geq \frac{1}{\langle x_1 \rangle} \frac{1}{1+Q}$, so the minimum CV is inversely proportional to the square root of N_2, in contrast to $\eta_{11} \geq \frac{1}{\langle x_1 \rangle}$ implied by the original Lestas *et. al.* bound, saying that minimum CV cannot be sub-Poisson. The above observation is consistent with example (6) in Section III-B, where small n corresponds to a N_2 comparable with $\frac{N_2}{N_1}$, and large n corresponds to a N_2 dominated by $\frac{N_2}{N_1}$.

In summary, from our lower bound (12) on noise suppression for coupled reactions, we see that coupling is most effective when τ_1 is not too small and when absolute signaling rate N_2 is at least comparable with relative signaling rate $\frac{N_2}{N_1}$. This suggests that coupling is most useful for energy efficient noise suppression, where $\frac{N_2}{N_1}$ is not too large, so utilizing N_2 through coupling to reduce noise further could be rather significant.

D. Coupling in two-component CRN

Here, in order to discover design principles for noise suppression with coupling, we study the achievable lower bound for noise in two-node chemical reaction networks using LNA, explicitly allowing coupling among reactions.

We start from the following formula for noise for stable two-node chemical reaction networks using LNA (for derivation see Section V-B):

$$\begin{aligned}\eta_{11} \geq & \frac{1}{H_{11}\langle x_1 \rangle} \min \Bigg\{ \langle s_{1|1} \rangle, \\ & \frac{1}{1+K} \left(\langle s_{1|1} \rangle - 2k \left| \langle s_{1|2} \rangle \right| + k^2 \frac{\langle s_{2|2} \rangle}{N} \right) \Bigg\}.\end{aligned} \tag{13}$$

Here $N = \frac{N_2}{N_1} = \frac{\langle R_2^{\pm} \rangle \tau_1}{\langle x_1 \rangle}$, with N_1,N_2 being the signaling rates as in the Lestas *et. al.* bound in equation (9), $k = \frac{H_{12}}{H_{22}} \operatorname{sgn}\left\{ \langle s_{1|2} \rangle \right\}$, and $K = -\frac{H_{12}H_{21}}{H_{11}H_{22}}$. k is the controller actuation gain, as H_{12} is the logarithmic gain of x_2's influence on x_1's birth and death rates. The sign of $\langle s_{1|2} \rangle$ enters k so that the sign of k is meaningful as well: $k < 0$ always implies amplification of noise, while $k > 0$ may result in suppression of noise. K is the closed loop gain, with negative sign in front to again make positive K correspond to noise suppression and negative K, noise amplification.

The two terms in the minimization in equation (13) are achievable (see Section V-A). η_{11} achieves the first term if the controller time scale is much larger than the plant time scale with $\tau_1 \gg \tau_2$, while the second term is achieved when the controller time scale is much faster than the plant time scale with $\tau_1 \ll \tau_2$. It is reasonable that η_{11} becomes $\eta_{11,nc} \equiv \frac{\langle s_{1|1} \rangle}{H_{11}\langle x_1 \rangle}$ when the controller x_2's dynamics is very slow, as x_2 is essentially constant in x_1's time scale, so x_1 is doing birth-death by itself with constant x_2. Therefore, we can regard $\eta_{11,nc}$, the first term, as a baseline that is easily achieved when there is no control at all. Note that the no control case is not the same as open loop, as open loop could have controller actuation on the plant with no sensing.

The more interesting term is the second one, which could become less than the first no-control term in certain architectures. For the second term to be small, we need $k > 0$ and $K > 0$. $K > 0$ corresponds to negative feedback, with either x_1 activating x_2 and x_2 repressing x_1, or x_1 repressing x_2 and x_2 activating x_1. $k > 0$ corresponds to an actuation that is compatible with the coupling reaction. If we have positive coupling, i.e. $\langle s_{1|2} \rangle > 0$, then $k > 0$ implies $H_{12} > 0$ or that the controller x_2 represses x_1, as in example (4) of Section III-A. If we have negative coupling, i.e. $\langle s_{1|2} \rangle < 0$, then $k > 0$ implies $H_{12} < 0$ or that the controller activates x_1, as in example (6) in Section III-B. More detailed implications are discussed below.

1) Feedback vs. feedforward: Both incoherent feedforward and negative feedback architectures are widely found in natural biological regulatory systems with homeostasis [29]. However, our result shows that feedforward architectures always have larger noise compared to negative feedback architectures with the same plant dynamics, controller actuation, and coupling. Feedforward architectures allow controller actuation but no sensing, so $K = 0$. If $K > 0$, which is the case for negative feedback, then the lower bound for noise is always lower than the feedforward case by a factor of $\frac{1}{1+K}$. Note that noise could be amplified by positive feedback with $K < 0$.

This is in strong contrast with the result for deterministic robust perfect adaptation, relevant for suppression of extrinsic noise (i.e. noise of parameters). Robust perfect adaptation is the property that a deterministic system could reach a homeostasis at steady state despite constant disturbances and uncertain dynamics [30], which is biologists' term for robust constant disturbance rejection. Both feedforward and feedback architectures could implement robust perfect adaptation equally well, with only differences in the physical controller implementation considerations [31]. Once noise is considered, we see that negative feedback could achieve smaller noise than feedforward, even when the underlying deterministic system dynamics is the same.

2) Coupled vs. decoupled.: The coupled case always have lower noise than the decoupled case if $k > 0$. If we assume our system does not have coupling, then $\langle s_{1|2} \rangle = 0$, so the term $-2k \left| \langle s_{1|2} \rangle \right|$, which is negative when $k > 0$, becomes zero for the decoupled case.

We also see that larger N or higher signaling rate always corresponds to less noise in the decoupled case, which agrees with what is implied by the Lestas *et. al.* bound. The lower bound for this case is $\eta_{11} \geq \eta_{11,fb} \equiv \frac{\langle s_{1|1} \rangle}{H_{11}(1+K)\langle x_1 \rangle}$, achieved in the large signaling ($N \to \infty$) and fast controller ($\tau_2 \ll \tau_1$) limit. Therefore we can regard $\eta_{11,fb}$ as the lowest noise achievable with only feedback regulations and no coupling. Note that $\eta_{11,fb} = \frac{1}{1+K}\eta_{11,nc}$, so we decrease noise by a factor of $\frac{1}{1+K}$ compared to no control case when we have negative feedback.

In comparison, the coupled case have a different interpretation of the relative signaling rate N, as $N = \frac{\langle s_{2|1} \rangle}{\langle s_{1|2} \rangle}$ if $\langle s_{1|2} \rangle \neq 0$, a consequence of equation (3). In this case, the relative signaling rate and the coupling strength are strongly related to each other. In particular, if the coupling reaction is

known, then the optimal signaling rate is no longer infinite. This is shown explicitly in the following subsection.

3) Limit on noise suppression by coupling: We show that the relative signaling rate N is not larger the better for coupled systems and that the suppression of noise through one coupled reaction is limited to a factor of $\frac{1}{2}$ beyond the feedback lower bound $\eta_{11,fb}$.

Consider the case where we have only one coupling reaction of the form $\{x_1, x_2\} \to \{x_1 + b_1\sigma_1, x_2 + b_2\sigma_2\}$, where $b_1, b_2 \in \mathbb{Z}_{>0}$ are the molecule counts of x_1 and x_2 changed by this reaction, and $\sigma_1, \sigma_2 \in \{1, -1\}$ are signs of the change. We assume this coupled reaction has α_1 portion of the birth or death flux of x_1, and α_2 portion of the flux of x_2, where $\alpha_1, \alpha_2 \in (0, 1]$. We also assume that all reactions other than the coupled one have step size 1. In terms of these parameters, we have $\langle s_{1|1} \rangle = \frac{\alpha_1 b_1 + (1-\alpha_1)+1}{2} = 1 + \frac{\alpha_1(b_1-1)}{2}$, $\langle s_{2|2} \rangle = 1 + \frac{\alpha_2(b_2-1)}{2}$, $\langle s_{1|2} \rangle = \frac{\alpha_2 b_1}{2}\sigma_1\sigma_2$, $\langle s_{2|1} \rangle = \frac{\alpha_1 b_2}{2}\sigma_1\sigma_2$, and $N = \frac{\langle s_{2|1} \rangle}{\langle s_{1|2} \rangle} = \frac{\alpha_1 b_2}{\alpha_2 b_1}$. Substituting these into equation (13) yields lower bounds for noise in terms of α_i and b_i. Notice that once the coupling reaction is known or b_1, b_2 are determined, varying the relative signaling rate N is the same as varying α_1 and α_2, the fractions of fluxes that the coupled reaction take.

The partial derivative of the second term of the lower bound in equation (13) with respect to b_1 is always positive, so b_1 should be as small as possible, while the partial derivative with respect to b_2 is always negative, so b_2 should be as large as possible. This makes sense, as larger $\frac{b_2}{b_1}$ means that x_1's change is amplified in x_2's change, so the signal is less corruptible by internal noise of x_2.

If we take $b_1 = 1$, then the partial derivative with respect to α_1 is always negative, so $\alpha_1 = 1$ is optimal. This is reasonable, as $\alpha_1 = 1$ means the coupling reaction accounts for as large a fraction of change as possible of x_1, so the knowledge of x_1's change obtained from x_2 through the coupling reaction is more informative.

The partial derivative with respect to α_2, however, is usually not as simple. The optimal α_2 is $\frac{\frac{\alpha_1 b_2}{k} - 1}{b_2 - 1}$ when $b_2 \neq 1$. This shows that the optimal fraction of flux in x_2 of the coupled reaction is related to the controller actuation gain as well as the stoichiometry of the coupled reaction. In particular, this shows that the optimal relative signaling rate N is determined by parameters of the coupled reaction, such as b_1, b_2 and k. For example, taking $\alpha_1 = 1$ and large b_2, we have $\frac{1}{k}$ as the optimal α_2, and the optimal effective signaling rate N is $k\frac{b_1}{b_2}$.

It should be noted that the relative signaling rate is not larger the better when the coupled reaction is fixed does not contradict that larger relative signaling rate should always reduce noise as indicated in equation (13). Indeed, larger effective signaling rate N always corresponds to smaller lower bound for noise if we do not constrain b_1 and b_2. The optimal b_2 is infinity, which corresponds to infinite N. However, once b_1 and b_2 are fixed, then N is no longer free to vary by itself. For example, if we fix parameters at their optimal values so that $b_1 = 1$, $\alpha_1 = 1$ and b_2 is large, then $N = \frac{1}{\alpha_2 b_2}$, which could only become larger by decreasing α_2. However, smaller α_2 would also mean that the coupling reaction accounts for a smaller fraction of x_2's change, so the signal about x_1 becomes more easily obfuscated by noise of x_2's other reactions not coupled with x_1. In other words, since smaller α_2 corresponds to more signal amplification through x_2's own birth-death reactions that are not coupled with x_1, there is a tradeoff between noisy amplification of signals through the decoupled reactions and the preservation of accurate signals through the coupled reactions.

Another implication of the above observation is that with coupled reactions, a moderate signaling rate could result in significant noise suppression, as the optimal N could be small. This suggests that noise suppression with coupling could be rather cost-effective, therefore preferable for biological systems.

Now, taking the optimal values of $b_1 = 1$, $\alpha_1 = 1$, $b_2 \to \infty$, and $\alpha_2 = \frac{1}{k}$ results in $\eta_{11} = \frac{1}{2}\eta_{11,fb}$ in the fast controller limit ($\tau_2 \ll \tau_1$), so the optimal lower bound with one coupling reaction while other reactions have step size 1 is half of the optimal lower bound of negative feedback. This shows that coupling cannot make noise arbitrarily small beyond the decoupled case.

IV. Discussion

In this work we investigated the coupling mechanisms that is ubiquitous in biology for noise suppression. We constructed examples that can suppress noise below the Lestas *et. al.* bound [2] and, by systematically analyzing all two-node chemical reaction networks, showed that coupling has implications on effective network architectures and that coupling has its own limitations in noise suppression.

Our example constructed in Section III-B suppresses noise below the Lestas *et. al.* bound, but broke one of their assumptions that require the plant and the controller to have separate birth-death reactions. It may therefore seem natural that we could go below their bound. However, it should be noted that not only is the Lestas *et. al.* bound hypothesized to be true in much more general situations [2], previous efforts in going beyond this bound either employed complex non-biological controllers [32] or could only show a theoretically better lower bound with unknown achievability [33]. Hence, finding an example or a class of systems that suppresses noise beyond the Lestas *et. al.* bound is highly nontrivial with significant implications. In particular, the example in Section III-B shows that biologically plausible reaction systems could indeed suppress noise below the Lestas *et. al.* bound, and situations beyond the bound should be seriously explored and considered for regulation and information processing in biological systems.

With that said, the results in this work does not try to contradict the results in [2], but rather follows its pioneering efforts in exploring fundamental limits of noise suppression in biological systems. We showed that noise suppression with coupling could further suppress noise, but still has its own fundamental limitations. This is also connected with the work of [3], which does not assume LNA and derives tight exact

lower bounds for noise, but the lower bounds are harder to analyze for different cases. It is therefore of high interest to derive exact lower bounds in the spirit of [3] that includes coupling. Furthermore, theoretical investigations that relate LNA analysis with the methods of [3] would be highly beneficial, as the LNA method is scalable and analytically tractable, hence useful for controller design and architecture exploration.

This work focused on intrinsic noise, while noise in biology could be intrinsic as well as extrinsic. Intrinsic noise arises from stochastic nature of the system itself, while extrinsic noise comes from fluctuations in system parameters [34]. Extrinsic noise could be treated rather well by ignoring intrinsic noise [34], so perfect extrinsic noise suppression could be considered as perfect adaptation in the deterministic system. Robust perfect adaptation and its constraints on system architecture for biomolecular systems is studied in [31]. It is therefore desirable to connect intrinsic noise suppression together with robust perfect adaptation concerns. After all, noise suppression is not meaningful if the fixed point is not desirable.

While coupling considered in this work is of a binary kind, where two components are either coupled or decoupled, several biological phenomena observed in nature also have a "soft" coupling interpretation. For example, genes that are spatially close to each other due to their position on the genome or its 3D structure are more likely to be transcribed together in a short period of time [8]. Indeed, the positioning of genes are evolutionarily selected and the 3D structure of the genome is highly regulated [12]. This urges the study of a generalization of the coupling notion here to include these cases. It is also worth investigating how these general "soft" coupling notions relate to the cases investigated in Section III-D where coupling is only a fraction of the birth-death fluxes.

Lastly, coupling is a phenomenon rather specific to biomolecular control. Although covariance is considered in canonical stochastic control theory, coupling is a structually and physically different way of influencing variables' covariance than regulations through birth and death rates. This echoes the results in [31], where physical constraints specific to biomolecular reaction systems is shown to make integral feedback implementation a nontrivial design problem in biology. With these observations, it is highly desirable to develop theoretical tools custom-fit for biological systems so as to gain insight for biomolecular control problems.

Acknowledgement

The authors would like to thank Noah Olsman for constructive discussions. F. X., J. Y. and J. C. D. are partially funded by the Defense Advanced Research Projects Agency (Agreement HR0011-16-2-0049). J. Y. acknowledges the support of NSF/Simons Foundation. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[2] I. Lestas, G. Vinnicombe, and J. Paulsson, "Fundamental limits on the suppression of molecular fluctuations," *Nature*, vol. 467, no. 7312, p. 174, 2010.

[3] A. Hilfinger, T. M. Norman, G. Vinnicombe, and J. Paulsson, "Constraints on fluctuations in sparsely characterized biological systems," *Phys. Rev. Lett.*, vol. 116, p. 058101, Feb 2016.

[4] N. Barkai and B.-Z. Shilo, "Variability and robustness in biomolecular systems," *Molecular Cell*, vol. 28, pp. 755–760, 2007.

[5] C. L. Partch, C. B. Green, and J. S. Takahashi, "Molecular architecture of the mammalian circadian clock," *Trends in Cell Biology*, vol. 24, pp. 90–99, 2013.

[6] J. Gaucher, E. Montellier, and P. Sassone-Corsi1, "Molecular cogs: Interplay between circadian clock and cell cycle," *Trends in Cell Biology*, vol. 28, pp. 368–379, 2018.

[7] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, "Probing the limits to positional information," *Cell*, vol. 130, pp. 153–164, 2007.

[8] B. A. et. al., *Molecular Biology of the Cell*, 6th ed. W. W. Norton Company, 2014.

[9] J. M. Schmiedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Blthgen, D. S. Marks, and A. van Oudenaarden, "MicroRNA control of protein expression noise," *Science*, vol. 348, pp. 128–132, 2015.

[10] N. Delalez, A. Sootla, G. H. Wadhams, and A. Papachristodoulou, "Design of a synthetic srna-based feedback filter module," *bioRxiv*, 2018. [Online]. Available: https://www.biorxiv.org/content/early/2018/12/21/504449

[11] F. A. Chandra, G. Buzi, and J. C. Doyle, "Glycolytic oscillations and limits on robust efficiency," *science*, vol. 333, no. 6039, pp. 187–192, 2011.

[12] G. Kungulovski1 and A. Jeltsch, "Epigenome editing: State of the art, concepts, and perspectives," *Trends in Genetics*, vol. 32, pp. 101–113, 2016.

[13] M. Adli, "The CRISPR tool kit for genome editing and beyond," *Nature Communications*, vol. 9, no. 1911, 2018.

[14] P. Bokes, M. Hojcka, and A. Singh, "Buffering gene expression noise by microRNA based feedforward regulation," in *Computational Methods in Systems Biology*, M. Češka and D. Šafránek, Eds. Cham: Springer International Publishing, 2018, pp. 129–145.

[15] G. Buzi and M. Khammash, "Implementation considerations, not topological differences, are the main determinants of noise suppression properties in feedback and incoherent feedforward circuits," *PLOS Computational Biology*, vol. 12, no. e1004958, 2016.

[16] L. Laurenti, A. Csikasz-Nagy, M. Kwiatkowska, and L. Cardelli, "Molecular filters for noise reduction," *Biophysical Journal*, vol. 114, no. 12, pp. 3000 – 3011, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S000634951830585X

[17] J. Paulsson, "Models of stochastic gene expression," *Physics of Life Reviews*, vol. 2, pp. 157–175, 2005.

[18] I. Lestas, J. Paulsson, N. E. Ross, and G. Vinnicombe, "Noise in gene regulatory networks," *IEEE Transactions on Automatic Control*, vol. 53, no. Special Issue, pp. 189–200, 2008.

[19] E. O. Volt, H. A. Martens, and S. W. Omholt, "150 years of the mass action law," *PLOS Computational Biology*, vol. 11, no. 1, 2015.

[20] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, p. 052150, 1992.

[21] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump markov processes," *Journal of Applied Probability*, vol. 7, pp. 49–58, 1970.

[22] D. Schnoerr, G. Sanguinetti, and R. Grima, "Approximation and inference methods for stochastic biochemical kinetics a tutorial review," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 093001, 2017.

[23] R. Zwanzig, *Nonequilibrium statistical mechanics*. Oxford University Press, 2001.

[24] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback control theory*. Courier Corporation, 2013.

[25] N. Olsman, F. Xiao, and J. C. Doyle, "Architectural principles for characterizing the performance of antithetic integral feedback networks," *iScience*, Accepted.

[26] N. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Amsterdam: Elsevier, 2007.

[27] C. Cianci, D. Schnoerr, A. Piehler, and R. Grima, "An alternative route to the system-size expansion," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 395003, 2017.
[28] J. E. Ferrell, "Perfect and near-perfect adaptation in cell signaling," *Cell Systems*, vol. 2, no. 2, pp. 62–67, 2016.
[29] U. Alon, *An Introduction to Systems Biology, Design Principles of Biological Circuits*. London: CRC, 2006.
[30] T.-M. Yi, Y. Huang, M. I. Simon, , and J. Doyle, "Robust perfect adaptation in bacterial chemotaxis through integral feedback control," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, 2000.
[31] P. A. Iglesias and A. Levchenko, "Robust perfect adaptation in biomolecular reaction networks," in *Proceedings of the 57th IEEE Conference on Decision and Control*, in-press.
[32] K. Parag and G. Vinnicombe, "Single event molecular signalling for estimation and control," in *Proceedings of the 2013 European Control Conference (ECC)*, vol. 1, 2013, pp. 4166–4171.
[33] Y. Nakahira, F. Xiao, V. Kostina, and J. C. Doyle, "Fundamental limits and achievable performance in biomolecular control," in *2018 Annual American Control Conference (ACC)*, June 2018, pp. 2707–2714.
[34] A. Hilfinger and J. Paulsson, "Separating intrinsic from extrinsic fluctuations in dynamic biological systems," *Proceedings of National Aacademy of Science*, vol. 108, pp. 12 167–12 172, 2011.
[35] R. S. Lipster and A. N. Shiryayev, *Statistics of random processes II: Applications*. Springer-Verlag, 1974.
[36] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Eidgenossiche Technische Hochschule Zurich, 1998.
[37] T. R. Gingrich, G. M. Rotskoff, and J. M. Horowitz, "Inferring dissipation from current fluctuations," *J. Phys. A: Math. Theor.*, vol. 50, no. 18, p. 184004, apr 2017.
[38] A. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 5–25, 1974.

V. Appendix

A. Two component CRN noise analysis

Here we derive equation (13) using LNA.

Equation (2) for the two dimensional ($n = 2$) case gives rise to the following system of equations:

$$\begin{bmatrix} 2\left(\frac{H_{11}}{\tau_1}\eta_{11} + \frac{H_{12}}{\tau_1}\eta_{12}\right) & \left(\frac{H_{11}}{\tau_1} + \frac{H_{22}}{\tau_2}\right)\eta_{12} + \frac{H_{12}}{\tau_1}\eta_{22} + \frac{H_{21}}{\tau_2}\eta_{11} \\ & 2\left(\frac{H_{21}}{\tau_2}\eta_{12} + \frac{H_{22}}{\tau_2}\eta_{22}\right) \end{bmatrix} = \begin{bmatrix} \frac{2\langle s_{1|1}\rangle}{\tau_1\langle x_1\rangle} & \frac{\langle s_{1|2}\rangle}{\tau_2\langle x_1\rangle} + \frac{\langle s_{2|1}\rangle}{\tau_1\langle x_2\rangle} \\ & \frac{2\langle s_{2|2}\rangle}{\tau_2\langle x_2\rangle} \end{bmatrix}.$$

We make the following assumptions: the deterministic linearized system is stable, and $H_{11}, H_{22} \neq 0$. The first assumption is needed for the LNA to have finite variance. Since the deterministic system is stable if and only if $-\boldsymbol{M}$ is Hurwitz, for $n = 2$ case we have the following conditions for stable linearized system: $\frac{H_{11}}{\tau_1} + \frac{H_{22}}{\tau_2} > 0$ while $H_{11}H_{22} - H_{12}H_{21} > 0$. The second assumption is more biological. $H_{11} = 0$ implies $\partial_{x_1} R^- = \partial_{x_1} R^+$, so stability of x_1 relies on sensing and actuation of other system components. This is highly undesirable biologically, as constantly mutations could occur in the system that results in loss of function, breaking the sensing reactions such that other components become fixed and do not respond to changes of x_1, or breaking the actuation reactions such that other components' actions on x_1 become zero. In either case, x_1 dynamics could become unstable, implying disastrous consequences for the biological organism. In addition, the $H_{11} = 0$ or $H_{22} = 0$ cases are not biologically informative but simple to analyze, and can be easily shown that they satisfy the bound (13) as well.

With the above assumptions, we could then write the equation above into more informative parameters and rewrite it into the following linear system of equations:

$$\begin{bmatrix} 1 & k_{1|2}\overline{\alpha} & 0 \\ k_{2|1} & 1 & k_{1|2} \\ 0 & k_{2|1}\alpha & 1 \end{bmatrix} \begin{bmatrix} H_{11}\eta_{11} \\ (T_{21}H_{11} + H_{22})\eta_{12} \\ T_{21}H_{22}\eta_{22} \end{bmatrix} = \begin{bmatrix} \langle s_{1|1}\rangle \\ 2\langle s_{1|2}\rangle \\ \frac{\langle s_{2|2}\rangle}{N} \end{bmatrix} \frac{1}{\langle x_1\rangle},$$

where $T_{ij} = \frac{\tau_i}{\tau_j}$ is a ratio of time scales, $k_{i|j} = \frac{H_{ij}}{H_{jj}}$ is the logarithmic sensitivity of x_i to changes in x_j normalized by the sensitivity of x_j to its own changes, $N = \frac{\langle R_2\rangle\tau_1}{\langle x_1\rangle}$ is the relative signaling rate as defined in [2], and $\alpha = \frac{H_{11}\tau_2}{H_{11}\tau_2 + H_{22}\tau_1}$ with $\overline{\alpha} = 1 - \alpha$. Solving the system for η_{11} yields the following:

$$\eta_{11} = \frac{1}{H_{11}\langle x_1\rangle}\Bigg(\alpha\langle s_{1|1}\rangle + \overline{\alpha}\frac{\langle s_{1|1}\rangle - 2k_{2|1}\langle s_{2|1}\rangle + k_{2|1}^2\frac{\langle s_{2|2}\rangle}{N}}{1 - k_{1|2}k_{2|1}}\Bigg).$$

So we see that η_{11} can be expressed as a convex combination of two terms since $\alpha + \overline{\alpha} = 1$ and $\alpha \in [0, 1]$. Taking the min of the two terms then yield the lower bound shown in equation (13), with the first term achieved when $\alpha = 1$ or $\tau_2 \gg \tau_1$, and the second term achieved when $\alpha = 0$ or $\tau_2 \ll \tau_1$.

B. Birth-death interpretation of LNA

LNA was considered an interpretable but inaccurate approximation because of its derivation assuming large mean and Gaussian noise, which violates the non-negativity and discreteness of the variables [26], [22]. Here, through one simple example, we show that the Gaussian interpretation is not necessary, and the non-negativity and discreteness could be preserved. A systematic investigation is ongoing work to be published on another occasion.

Consider the following non-linear birth death process:

$$x \xrightarrow{a^3} x + 1 \qquad x \xrightarrow{x^3} x - 1.$$

Equation (2) gives $\eta = \frac{1}{H\langle x\rangle} = \frac{1}{3\langle x\rangle}$. The only fixed point is $x^* = a$. Then consider the following linear-propensity birth-death process:

$$x \xrightarrow{r_1(x)} x + 1, \qquad x \xrightarrow{r_2(x)} x - 1,$$

where $r_2(x) = \frac{a^3}{a - \lceil\frac{2}{3}a\rceil}\left(x - \lceil\frac{2}{3}a\rceil\right)$ and $r_1(x) = r_2(x) - 3a^2(x - a)$, for $a > 3$. Therefore $R^- - R^+$ for this system is the same as that for the original nonlinear birth-death process, and equation (2) would yield the same result $\eta = \frac{1}{3\langle x\rangle}$ with the same fixed point x^*. On the other hand, the system does not evolve below $\lceil\frac{2}{3}a\rceil$ as the death rate is zero on that number, so the variable x is always non-negative.

This implies that LNA could be a good approximation even when the mean molecular count is small (in fact as small as 3 in this case), and the distribution could preserve the discreteness as well as non-negativity, so it is not necessarily Gaussian. The only situations that LNA breaks down then are the same as the situations where the deterministic linearization of a nonlinear dynamical system breaks down: sharp non-lineariarities away from the fixed point.

C. Channel capacity of coupled reactions

The key step to derive the lower bound of noise suppression with or without coupling is the find the lower and upper bounds of the mutual information between x_1 and x_2. Here we first introduce the mathematical definition of mutual information between two random processes (rather than variables), and compute the channel capacity of a coupled reaction.

Consider a complete probability space $(\Omega, \mathscr{F}, \mathrm{P})$ with nondecreasing right continuous family of sub-σ algebras $(\mathscr{F}_t), t \in [0, T]$. Let $(n_t, \mathscr{F}_t)$ be a Poisson process with its intensity λ_t being a function of of the input signal $\theta = \{\theta_s, s \leqslant t\}$ and also $n = \{n_s, s \leqslant t\}$ in an non-anticipatory manner:

$$n_t \xrightarrow{\lambda_t(t,\theta,n)} n_t + 1 \ , \tag{14}$$

where the non-anticipatory constraint requires that $\lambda_t(t, \theta, n)$ is $\mathscr{F}_t$-measurable. If a certain coding $\lambda_t(t, \theta, n), 0 \leqslant t \leqslant T$ is given, then a natural question is how much information is contained in the received signal n about the transmitted signal θ. By definition, the mutual information is:

$$\mathcal{I}_T(n, \theta) = \mathbb{E} \ln \frac{d\mu^T_{n,\theta}}{d[\mu^T_n \times \mu^T_\theta]}(n, \theta) \tag{15}$$

where μ^T_n, μ^T_θ, and $\mu^T_{n,\theta}$ are measures corresponding to processes n_t, θ_t, and (n_t, θ_t), $0 \leqslant t \leqslant T$.

The Liptser-Shiryaev formula shows that [35]:

$$\mathcal{I}_T(n, \theta) = \mathbb{E} \int_0^T \big(\lambda_t \ln \lambda_t - \hat{\lambda}_t \ln \hat{\lambda}_t\big) dt, \qquad \hat{\lambda}_t = \mathbb{E}[\lambda_t | \mathscr{F}^n_t], \tag{16}$$

immediately we can see if λ_t is $\mathscr{F}^\theta_t$-measurable, then $\mathcal{I}_T(n, \theta) = 0$ since $\lambda_t = \hat{\lambda}_t$.

By the causality nature of such channel, the Liptser-Shiryaev formula is actually the formula for directed information: $\mathcal{I}_T(\theta \to n)$ (see [36] for a complete introduction of directed information). Here we show that if we have one another parallel channel:

$$n_t \xrightarrow{\beta_t(t,\theta,n)} n_t \pm 1 \ , \tag{17}$$

where $\beta_t(t, \theta, n)$ is again $\mathscr{F}_t$-measurable, then the mutual information (or more precisely, the directed information) from θ_t to n_t is the sum of mutual (directed) information of each channel:

$$\mathcal{I}_T(\theta \to n) = \mathbb{E} \int_0^T \Big(\lambda_t \ln \frac{\lambda_t}{\hat{\lambda}_t} + \beta_t \ln \frac{\beta_t}{\hat{\beta}_t}\Big) dt \tag{18}$$

Equation (18) can be bounded by using Lagrangian multiplier methods, assuming the first and second moments of the propensities are finite, therefore we obtain an upper bound of the mutual information of two parallel Poisson channels (equation (14) and (17)) [2]:

$$\frac{1}{T}\mathcal{I}_T(\theta, n) \leq \langle\lambda\rangle \ln \left(1 + \frac{\mathrm{Var}\,\{\lambda\}}{\langle\lambda\rangle^2}\right) + \langle\beta\rangle \ln \left(1 + \frac{\mathrm{Var}\,\{\beta\}}{\langle\beta\rangle^2}\right) \tag{19}$$

Now let's turn to the coupled reaction, using equation (11) as an example. Equation (11) is equivalent to two Poisson channels:

$$x_1 \xrightarrow{\lambda_t[\mathcal{I}_t(x_1,x_2)]} x_1 - 1, \qquad x_2 \xrightarrow{w(x_1)} x_2 + 1 \tag{20}$$

where λ switches between 0 and a large constants K. It switches from 0 to K when x_2 increases by 1, then switches back to 0 when x_1 also increases by 1. Hence the overall effect of the first reaction to tracking the dynamics of the second reaction, when the second reaction occurs, the first immediately occurs, when setting $K \to \infty$. Then the channel capacity of this coupled reaction can be calculated by summing up the channel capacity of the two channels (equation (20)). The second one, as the same as a usual Poisson channel, has a channel capacity $C_2 = \mathrm{Var}\,\{w(x_1\}\,/\,\langle w(x_1)\rangle$. For the first reaction, since λ basically tracks the number of occurrence of the second one, we instead can use a classic conclusion from large deviation theory, which claims that $\mathrm{Var}\,\{\lambda\}$ is equal to its mean [37]. On the other hand, in the limit of long time, the empirical mean of number of events is equal to $\langle w(x_1)\rangle$, then from ergodic theorem we conclude that $\langle\lambda\rangle = \langle w(x_1)\rangle$. Therefore the channel capacity of the coupled reaction is then:

$$C = \frac{\mathrm{Var}\,\{w(x_1)\}}{\langle w(x_1)\rangle} + \langle w(x_1)\rangle \ln(1 + \frac{1}{\langle w(x_1)\rangle}) \tag{21}$$

Furthermore, coupling does not alter the lower bound on mutual information, which is derived from Pinsker's non-anticipatory ε-entropy [33], [38]. So the lower bound on noise suppression is only due to an increase of channel capacity.