bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Antithetic Integral Feedback: A new motif for robust perfect adaptation in noisy biomolecular networks

Corentin Briat[†], Ankit Gupta[†], Mustafa Khammash[*]
Department of Biosystems Science and Engineering (D-BSSE),
ETH–Zürich, Basel, Switzerland

November 23, 2015

Abstract

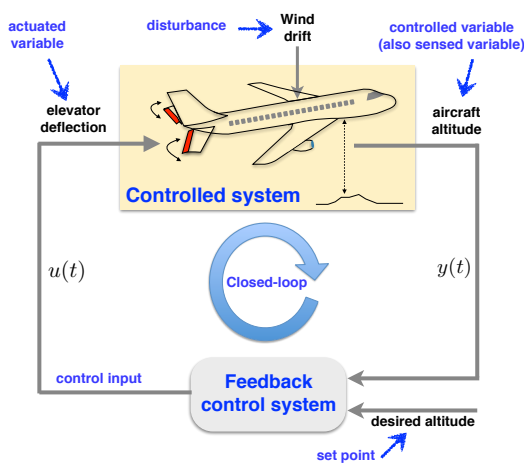
Homeostasis is a running theme in biology. Often achieved through feedback regulation strategies, homeostasis allows living cells to control their internal environment as a means for surviving changing and unfavourable environments. While many endogenous homeostatic motifs have been studied in living cells, some other motifs may remain under-explored or even undiscovered. At the same time, known regulatory motifs have been mostly analyzed at the deterministic level, and the effect of noise on their regulatory function has received low attention. Here we lay the foundation for a regulation theory at the molecular level that explicitly takes into account the noisy nature of biochemical reactions and provides novel tools for the analysis and design of robust homeostatic circuits. Using these ideas, we propose a new regulation motif, which we refer to as *antithetic integral feedback*, and demonstrate its effectiveness as a strategy for generically regulating a wide class of reaction networks. By combining tools from probability and control theory, we show that the proposed motif preserves the stability of the overall network, steers the population of any regulated species to a desired set point, and achieves robust perfect adaptation – all with low prior knowledge of reaction rates. Moreover, our proposed regulatory motif can be implemented using a very small number of molecules and hence has a negligible metabolic load. Strikingly, the regulatory motif exploits stochastic noise, leading to enhanced regulation in scenarios where noise-free implementations result in dysregulation. Finally, we discuss the possible manifestation of the proposed antithetic integral feedback motif in endogenous biological circuits and its realization in synthetic circuits.

Perfect adaptation is that property of a biological system (e.g. a cell) which enables it to adapt to an external stimulus so that it maintains responsiveness to further stimuli. To be effective, such an adaptation mechanism must be robust, i.e. it must remain functional over a wide range of stimulus levels and system parameters. It was shown in [29] that robust perfect adaptation in bacterial chemotaxis is achieved due to integral feedback control in the prevalent chemotaxis model [3]. Other homeostatic systems have also been shown to realize integral feedback control. For example it was demonstrated in [8] that calcium homeostasis in mammals relies on an integral feedback strategy to achieve perfect adaptation to persistent changes in plasma calcium clearance or influx, a property that enables mammals to maintain physiological levels of plasma calcium within tight tolerances in spite of varying demands for calcium. In [19], integral feedback was implicated in the robust regulation of membrane turgor pressure in Saccharomyces cerevisiae. Following an osmotic shock, nuclear enrichment of the MAP kinase Hog1 adapts perfectly to changes in external osmolarity, a result of an integral feedback action that requires Hog1 kinase activity. However, as some theoretical studies have suggested [14, 23], adaptation may not be solely related to integral control .

[*]Corresponding author (e-mail: mustafa.khammash@bsse.ethz.ch); [†]Authors contributed equally.

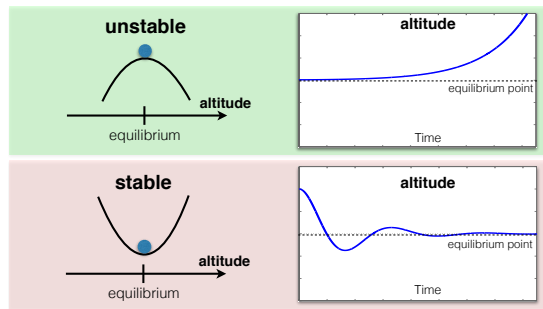
bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Elements of Feedback Control Systems. Feedback control systems are workhorses of modern technology. They are also prevalent at every level of biological organization. *Control theory* deals with the analysis and design of such systems. We outline the key elements of feedback control systems using a familiar example–a modern airplane. The goal of every feedback control system is to control (i.e. regulate) the behavior of a given dynamical system (**the controlled system**). This is achieved by manipulating an input (**actuated variable**) to affect the system dynamics in a way that the system output (**controlled variable**) follow a certain behavior. In the case of our airplane example, the input variable is the deflection angle of the elevator, while the regulated variable is the airplane altitude. The control objective is to reach a desired altitude (**set point**) and to remain there over time, in spite of external factors like wind drift (**disturbances**) that tend to drive the system into disequilibrium.



Maintaining set point tracking is the task of the feedback control system which uses continuous measurements of a dynamic variable (**sensed variable**) together with the desired set point and prior information about the system dynamics to generate a correction signal (**control input**) that is fed back to the dynamical system as its actuated variable, thereby 'closing the loop'. The control system has its own dynamics and together with the dynamics of the controlled system they form the aggregate dynamics of the closed-loop system, which has new properties that reflect the tight coupling of both.

When coupling the dynamics of two systems through feedback, one of the key requirements is the dynamic stability of the new closed loop system (**closed-loop stability**). The stability of each system in isolation does not guarantee the stability of the closed-loop, and one must carefully design the control system to achieve closed-loop stability. Once stability is assured, one can begin to



explore the robust set point tracking property of the system. Mathematically, set point tracking is achieved when: $e(t) \to 0$ as $t \to \infty$, where $e(t)$ is the error signal, defined as the difference between the controlled output $y(t)$ and the desired set point, r. The above condition assures that the controlled variable tracks the desired set point after some time has elapsed. Achieving such tracking performance in spite of unknown disturbances and unknown dynamics of the controlled system is referred to as *robust tracking* and is an important goal of many control systems. Control theory stipulates that control systems with certain structural properties, such as integral feedback action, will achieve set point tracking, and that tracking will be maintained regardless of system parameters and constant disturbances (provided stability is preserved).

Dynamics:

$\dot{y}(t) = f(y(t), u(t))$ *controlled dynamical system*

$\dot{u}(t) = g(u(t), y(t), r)$ *feedback control system*

Objectives:

1. **Closed-loop stability**
2. **Robust tracking:** $e(t) \to 0, \quad where \;\; e(t) := r - y(t)$
   (in spite of disturbances and unknown system parameters)

Possible Control Strategies:

- **Proportional-Integral feedback:** $u(t) = K_P e(t) + K_I \int_0^t e(\tau) d\tau$
- **Integral feedback:** $K_P = 0$
- **Proportional feedback:** $K_I = 0$

Example:

Dynamical system (no feedback control)

$\dot{y}(t) = \theta y(t) + u(t)$ (unstable dynamics for $\theta \geq 0$)

Closed-loop dynamics (with feedback control)

with proportional feedback: $\dot{y}(t) + (K_P - \theta) y(t) = K_P r$

stable dynamics when $K_P > \theta$, and $y(t) \to K_P r/(K_P - \theta)$ no tracking

with integral feedback: $\ddot{y}(t) - \theta \dot{y}(t) + K_I y(t) = K_I r$

unstable dynamics if $\theta \geq 0$.

stable dynamics and $y(t) \to r$ (provided $\theta < 0$ and $K_P > 0$) *robust tracking!*

with proportional-integral feedback: $\ddot{y}(t) + (K_p - \theta)\dot{y}(t) + K_I y(t) = K_I r$

stable dynamics and $y(t) \to r$ (provided $K_P > \theta$ and $K_I > 0$) *robust tracking!*

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In engineering applications, integral feedback is recognized as a principal strategy for regulation. The Proportional-Integral-Derivative (PID) control architecture, which includes integral feedback as an essential element, is the workhorse of industrial control and is implemented in the majority of all automatic control applications [1]. Undoubtedly, the prevalence of such a control strategy in natural and man-made systems is due to the inherent property of integral feedback control to robustly steer a regulated system variable to a desired set point, while achieving perfect adaptation to disturbances (or stimuli), regardless of the model parameters. Perhaps surprisingly, engineered biological circuits displaying perfect adaptation have received little attention so far, and current synthetic circuits only rely on simpler feedback strategies. For example, several control loops for controlling the level of biofuel production in bacteria while still maintaining a low toxicity level are theoretically analyzed in [7]. Another synthetic negative feedback loop is also designed in [24] for the control of protein translation. Instead of integral feedback strategies, these circuits rely on the simpler proportional feedback strategy. Consequently, they require a cumbersome tuning of parameters for achieving their goals. Such a tuning is very difficult to realize in a biological setting, and even if a proportional feedback strategy is perfectly implemented, the absence of integral action implies that it does not possess the key property of perfect adaptation.

In the noise-free (deterministic) setting, integral feedback control is well-understood, and its ability to achieve robust set-point tracking and perfect adaptation is well-known [1]. In contrast, analogous strategies in intrinsically noisy cellular environments are unknown. Indeed in biologically important settings where the dynamics is described by stochastic processes (e.g. continuous-time discrete-state Markov processes), determining what constitutes integral feedback remains unclear. As in the deterministic case, a "stochastic integral feedback" strategy must achieve closed-loop stability of the overall system, robust set-point tracking and robust perfect adaptation. Unlike the deterministic setting, however, set-point tracking and adaptation robustness must be maintained not only with respect to model parameters, but also for the highly fluctuating species abundances. One possible way to construct "stochastic integral feedback" is to use statistical moments (such as the mean or the variance) to describe the process to be regulated, and then to design feedback regulation strategies that steer these moments to desired values while achieving perfect adaptation [17]. While this approach brings the problem back to the deterministic domain (statistical moments evolve according to deterministic dynamics), one is immediately faced with the moment closure problem, whereby an infinite set of differential equations is needed to determine even the first two moments; see e.g. [11]. Similar difficulties arise if one works with the chemical master equation; see e.g. [9].

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

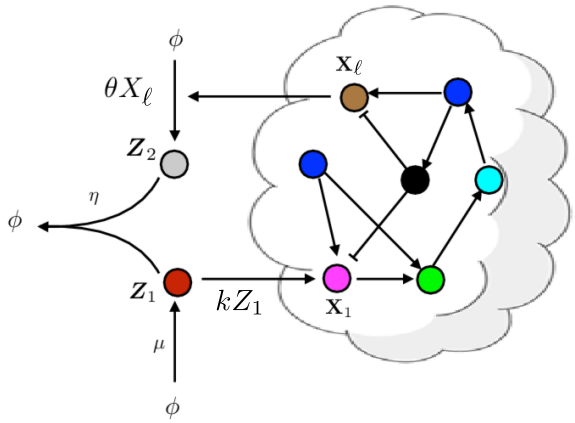


Figure 1: Schematic representation showing the constituents of our biomolecular control system. The network on the right (inside the cloud) represents the open-loop network, whose dynamics are to be controlled. Control is achieved by augmenting another network of reactions, referred to as the controller network (outside the cloud). Together the two networks form the closed-loop network whose dynamics are determined by the coupling resulting from the interaction of both networks. The proposed controller network, which we refer to as antithetic integral controller, acts on the open-loop network by influencing the rate of production of the actuated species $\boldsymbol{X_1}$ by means of the control input species $\boldsymbol{Z_1}$. The regulated species $\boldsymbol{X_\ell}$ will be influenced by the increase or decrease of the actuated species $\boldsymbol{X_1}$ and, in return, will influence the rate of production of the sensing species $\boldsymbol{Z_2}$, that will, finally, annihilate with the control input species $\boldsymbol{Z_1}$, thereby implementing a negative feedback control loop. The integral action is encoded in all the reactions of the controller network.

Here we adopt a novel approach for designing a stochastic integral feedback strategy that exhibits robust set-point tracking and robust perfect adaptation. Rather than dealing with the deterministic moments dynamics, we work with the stochastic chemical reaction network directly, thereby circumventing the moment closure problem. The objective of our control setup, represented in Fig. 1, is to bring the population average of the species $\boldsymbol{X_\ell}$ involved in a network (the "cloud" in Fig. 1) to a desired set-point. To achieve this, a new set of chemical reactions is introduced in a way that effectively implements a "stochastic antithetic integral feedback controller". This controller network consists of four reactions and two additional controller species ($\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$) that can annihilate each other. The species $\boldsymbol{Z_1}$ actuates the network which, in turn, influences the production of $\boldsymbol{Z_2}$ through the output species $\boldsymbol{X_\ell}$. We show in the results section that, for a large class of networks, the steady-state value for the population average of $\boldsymbol{X_\ell}$ depends exclusively on the ratio of two of the controller parameters, and is independent of the network parameters. In this respect, the closed-loop network exhibits stability, robust set-point tracking, and robust perfect adaptation for $\boldsymbol{X_\ell}$. To analyze such stochastic systems and to guarantee that they achieve these objectives, a new theory is needed. We develop such a theory here and use it to show that for a large class of networks, the considered antithetic feedback control motif can be used to achieve the desirable properties of "stochastic integral feedback". We rigorously prove that such a motif robustly achieves the desired closed-loop stability (ergodicity) property. We additionally show that it achieves robust set-point tracking and robust perfect adaptation under mild conditions on the uncontrolled network. Intriguingly, our " stochastic antithetic integral control motif" can provably achieve all the desired properties mentioned above, even when very low molecular copy numbers exist anywhere in the network. This presents a clear advantage in synthetic biology applications, where synthetic control loops involving large molecular counts can impose a debilitating metabolic load on the cell. Our control scheme can also be shown to possess remarkable stabilizing properties that are not found in deterministic implementations of the same circuit.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This provides a clear example where the intrinsic stochastic noise is beneficial–it stabilizes a system which would otherwise be unstable. To the best of our knowledge, such a beneficial effect of noise, in the context of control theory, is reported here for the first time. Note that many other benefits of noise, such as stochastic focusing [22], noise-induced oscillators [28] and noise-induced switches [2, 25], have appeared in the literature in recent years.

Results

In what follows, we elaborate on the control problem under consideration, the proposed controller, along with some technical results stating the conditions under which the proposed controller solves the considered control problem. Interestingly, these conditions obtained from probability theory connect to well-known concepts of control theory, such as stability and controllability. Some additional properties, such as robustness and innocuousness, are also discussed.

The Network to be Controlled (Open-Loop Network)

We start by describing the reaction network we aim to control. Consider a reaction network with mass-action kinetics involving d molecular species denoted by $\boldsymbol{X_1}, \ldots, \boldsymbol{X_d}$. Under the well-mixedness assumption [4], we can model the dynamics by a Markov process whose state at any given time is simply the vector of molecular counts of the d species. The "Markovian" assumption on the dynamics means that given the current state of the system, the future evolution of the state is independent of the past (memoryless property). The state evolution is influenced by K reaction channels: if the state is x, then the k-th reaction fires at rate $\lambda_k(x)$ and it displaces the state by the *stoichiometric vector*, $\zeta_k \in \mathbb{Z}^d$, where $\mathbb{Z}$ denotes the set of integers. Here λ_k is called the *propensity function* of the k-th reaction and is assumed to satisfy the property that if for any $x \in \mathbb{N}_0^d$ we have $x + \zeta_k \notin \mathbb{N}_0^d$, then $\lambda_k(x) = 0$, where $\mathbb{N}_0$ denotes the nonnegative integers. This property ensures that molecular counts of all the species remain nonnegative throughout the dynamics. In the following, we shall refer to this network as the *open-loop reaction network* and denote it by $(\boldsymbol{X}, \lambda, \zeta)$.

We now fix a state-space $\mathcal{S}$ for the Markovian reaction dynamics. This set $\mathcal{S}$ is a non-empty subset of $\mathbb{N}_0^d$ which is closed under the reaction dynamics. This means that for any state $x \in \mathcal{S}$ we must also have $(x + \zeta_k) \in \mathcal{S}$ if the the k-th reaction has a positive rate of firing ($\lambda_k(x) > 0$) at state x. Selecting $\mathcal{S}$ this way allows us to use it as a generic state-space for all Markov processes describing the reaction kinetics and starting at an initial state in $\mathcal{S}$ [10]. Henceforth, we denote by $\{X(t) = (X_1(t), \ldots, X_d(t)) : t \geq 0\}$ the continuous time Markov process representing the reaction dynamics with an initial state $x_0 \in \mathcal{S}$.

From a control theoretic point of view, it is necessary to define input and output nodes of the above network. We assume here that species $\boldsymbol{X_1}$ is the *actuated species* which is the species that the controller can act on. The *regulated species* is $\boldsymbol{X_\ell}$, for some $\ell \in \{1, \ldots, d\}$, and it is the species we wish to control. The way the controller acts on the actuated species, in order to control the regulated species is depicted in Fig. 1. This will be explained in more detail in the next section.

The Control Objectives

We now state the objectives required from our control system, which includes an open-loop network and a controller network.

Objectives. *Find a controller (set of additional reactions and additional species) such that, by suitably acting on the actuated species $\boldsymbol{X_1}$, we have the following properties for the closed-loop network (defined*

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

here as the interconnection of the open-loop network $(\boldsymbol{X}, \lambda, \zeta)$ *described above with the controller network):*

1. *the closed-loop network is ergodic;*
2. *the first and second-order moments of* $X(t)$ *exist and are uniformly bounded and globally converging with time to their unique stationary value;*
3. *we have that* $\mathbb{E}[X_\ell(t)] \to \mu^*$ *as* $t \to \infty$ *for some desired set-point* $\mu^* > 0$.

The first requirement is fairly standard. Indeed, ergodicity is the analogue of having a globally attracting fixed point for deterministic dynamics (i.e. global stability) and is required here so that the closed-loop network is well-behaved, in the sense that it reaches stationarity starting from any initial distribution. The second requirement is more specific to stochastic processes, as even if the means converge, the variance can still grow unboundedly with time, which would mean that the actual dynamics of the process (its sample-paths) is not well-behaved, rendering the controller of little practical utility. Finally, the third statement encapsulates our desired objective of perfect adaptation (or set-point tracking), i.e. that the population mean of the regulated species $\boldsymbol{X_\ell}$ approaches a fixed homeostatic value μ^*.

The Controller Reactions

We propose the following controller network (Fig. 1) inspired from the deterministic networks proposed in [21]:

$$\begin{aligned} &\underbrace{\emptyset \xrightarrow{\mu} \boldsymbol{Z_1}}_{\text{reference}}, && \underbrace{\emptyset \xrightarrow{\theta X_\ell} \boldsymbol{Z_2}}_{\text{measurement}}, \\ &\underbrace{\boldsymbol{Z_1} + \boldsymbol{Z_2} \xrightarrow{\eta} \emptyset}_{\text{comparison}}, && \underbrace{\emptyset \xrightarrow{kZ_1} \boldsymbol{X_1}}_{\text{actuation}}. \end{aligned} \tag{1}$$

where $\boldsymbol{X_\ell}$ is the species *measured* by the controller, which is identical to the *regulated species* in the current setup. The species $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are referred to as the *controller species*. Note that the topology of the controller network belongs to a family of four control topologies (see Fig. **??** to Fig. **??** in the Supplementary material) depending on respective roles of the two controller species.

Although inspired from [21], the above network has a different philosophy. Besides the fact that the current setting is stochastic, the main difference lies in the way the network interacts with the environment. While the goal of [21] was the biomolecular implementation of linear input-output systems, the goal here is to control a reaction network. In this regard, the birth-reactions of $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ clearly differ from the way they are defined in [21]. We now clarify the role and meaning of each of the controller reactions:

1. The first reaction is the *reference reaction* (or set-point) which (partially) sets the value of the reference $\mu^* = \mu/\theta$. This value is implemented as the birth-rate of species $\boldsymbol{Z_1}$.
2. The second reaction is the *measurement reaction* and takes the form of a pure-birth reaction with a rate proportional to the current population of the regulated species $\boldsymbol{X_\ell}$[1]. It is referred to as the *measurement reaction* as the rate of increase of the population of $\boldsymbol{Z_2}$ reflects the population size of $\boldsymbol{X_\ell}$.
3. The third reaction implements the *comparison reaction*, which decrements the molecular counts of $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ by one each. The rate constant for this reaction is η that can be tuned. The main role of

[1]Note that it can also be implemented in terms of the catalytic reaction $\boldsymbol{X_\ell} \xrightarrow{\theta} \boldsymbol{X_\ell} + \boldsymbol{Z_2}$.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

this reaction is to correlate both the populations of $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ and to prevent them from growing without bounds. This reaction can be viewed as a *compare and substract operation* since when both $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ have positive populations (comparison), then this reaction decreases their respective population-sizes by one (subtraction), thereby preserving the difference of their population-sizes.

4. The last reaction is the *actuation reaction*, which implements the way the controller acts on the system, i.e. by acting on the birth-rate of the actuated species $\boldsymbol{X_1}$[2]. The parameter k is also a tuning parameter of the controller.

The above controller has been chosen with an implementability constraint in mind, as it is expressed as plausible reactions that may be implemented in-vivo to perform *in-vivo control*. It will be shown later that the proposed controller exhibits strong robustness properties which make its implementation much easier than other types of controllers that require the fine tuning of their reaction rates (see also Section **??** of the supplementary material). In-vitro control is also possible using, for instance, DNA strand displacement [6]. In-silico control [17, 27] can be considered as well, whenever the population size of regulated species $\boldsymbol{X_\ell}$ can be measured in real-time from outside of the cell(s) using, for instance, microscopy.

Guaranteed Performance Properties of the Controlled Network

We now consider the dynamics of the *closed-loop network*, which is formed by interconnecting the open-loop network $(\boldsymbol{X}, \lambda, \zeta)$ with the controller (1). We can represent the dynamics by a Markov process $\{(X(t), Z(t)) : t \geq 0\}$, where for each t, $X(t) = (X_1(t), \ldots, X_d(t))$ denotes the molecular counts of the d network species and $Z(t) = (Z_1(t), Z_2(t))$ denotes the molecular counts of the two controller species. Our control objective is to steer the first-order moment $\mathbb{E}[X_\ell(t)]$ corresponding to species $\boldsymbol{X_\ell}$ to a desired set-point μ^*. The dynamics of the first-order moments $\mathbb{E}[X(t)] = (\mathbb{E}[X_1(t)], \ldots, \mathbb{E}[X_d(t)])$ and $\mathbb{E}[Z(t)] = (\mathbb{E}[Z_1(t)], \mathbb{E}[Z_2(t)])$ is described by the following system of ordinary differential equations (ODEs)

$$\begin{aligned} \frac{d\mathbb{E}[X(t)]}{dt} &= \sum_{k=1}^{K} \zeta_k \mathbb{E}[\lambda_k(X(t))] + k\mathbb{E}[Z_1(t)]e_1, \\ \frac{d\mathbb{E}[Z_1(t)]}{dt} &= \mu - \eta\mathbb{E}[Z_1(t)Z_2(t)], \\ \frac{d\mathbb{E}[Z_2(t)]}{dt} &= \theta\mathbb{E}[X_\ell(t)] - \eta\mathbb{E}[Z_1(t)Z_2(t)], \end{aligned} \tag{2}$$

where e_1 is the d-dimensional vector whose first component is 1 and the rest are zero. Note that this system of ODEs is not *closed* because there is no equation for the dynamics of $\mathbb{E}[Z_1(t)Z_2(t)]$. Moreover if the propensity functions λ_k's are nonlinear functions of the state-variables, additional quantities whose dynamics is not captured by these equations will be encountered. Attempting to "close" this system by adding equations for the dynamics of all these additional quantities and $\mathbb{E}[Z_1(t)Z_2(t)]$ will again lead to another set of quantities with unrepresented dynamics. This is known as the *moment-closure problem* in the literature, a well-known barrier for the direct analysis and simulation of moment equations.

As stated, it is unclear why the proposed controller structure involves an integral action. To emphasize this feature, let us define $\delta_Z(t) := Z_1(t) - Z_2(t)$. By subtracting the two last equations in (2), we get that

$$\frac{d\mathbb{E}[\delta_Z(t)]}{dt} = \mu - \theta\mathbb{E}[X_\ell(t)]. \tag{3}$$

[2]This can also be represented by the catalytic reaction $\boldsymbol{Z_1} \xrightarrow{k} \boldsymbol{Z_1} + \boldsymbol{X_1}$.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Defining the set-point *tracking error* as $e(t) := \mu/\theta - \mathbb{E}[X_\ell(t)]$ and integrating the above expression over the interval $[0, t]$ yields

$$\mathbb{E}[\delta_Z(t)] = \theta \int_0^t e(s)ds + \mathbb{E}[\delta_Z(0)], \tag{4}$$

which allows us to conclude that the first-order moment of the control input, $\mathbb{E}[u(t)] = k\mathbb{E}[Z_1(t)]$, will depend on the integral of the set-point tracking error $e(t)$ through the term $\mathbb{E}[\delta_Z(t)]$. Even though the control input only partially depends on the integrator state (4), the stability of the closed-loop system proved in the main results indicates that the integral action is preserved.

Our approach in this paper is to find conditions ensuring that the Markov process $\{(X(t), Z(t)) : t \geq 0\}$ describing the dynamics of closed-loop reaction network is ergodic. This means that starting from any initial state, the probability distribution of the state $(X(t), Z(t))$ converges to a unique stationary distribution π as $t \to \infty$. Under fairly general conditions, ergodicity also implies that first and second-order moments of the state $(X(t), Z(t))$ converge to their steady-state values, which can be computed by evaluating the expectation $\mathbb{E}_\pi$ with respect to the stationary distribution π. Furthermore, as $t \to \infty$, the right hand side of (3) tends to 0 and this yields the expression:

$$\mu - \theta\mathbb{E}_\pi[X_\ell] = 0 \tag{5}$$

where $\mathbb{E}_\pi[X_\ell] = \lim_{t\to\infty} \mathbb{E}[X_\ell(t)]$. From (5), we can immediately conclude that $\mathbb{E}_\pi[X_\ell] = \mu/\theta = \mu^*$, our desired objective. This shows that if the closed-loop network dynamics is ergodic, then our proposed controller automatically imposes the set-point tracking property, $\mathbb{E}[X_\ell(t)] \to \mu^*$ as $t \to \infty$, regardless of the initial conditions. Obtaining such a behavior is the main rationale behind integral control. The controller we propose performs integral action on the dynamics, and achieves set-point tracking and perfect adaptation properties. Note that we demonstrate these properties of our controller without solving the first-order moment equations (2), thereby circumventing the moment-closure problem mentioned above.

The following result, proved in Section **??** of the supplementary material, establishes conditions under which a general stochastic biochemical reaction network can be controlled using the proposed controller network.

Theorem 1 (General Network Case). *Consider an open-loop reaction network* $(\boldsymbol{X}, \lambda, \zeta)$ *and assume that for some given values of its parameters, and the parameters of the controller network, the closed-loop network, formed by augmenting the open-loop network with the controller reactions* (1)*, is ergodic and has uniformly bounded first- and second-order moments. Then, asymptotic set-point tracking is achieved, i.e.* $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$.

The main challenge in the above result lies in verifying the ergodicity of a given closed-loop network dynamics. In what follows, we provide simple conditions that allow us to check this property using efficient computational techniques, such as linear programming. This is done for two main classes of networks, namely those consisting of *unimolecular* reactions and those consisting of *bimolecular* reactions. In the unimolecular case, for each reaction k the propensity function $\lambda_k(x)$ is an affine function of the state variable $x = (x_1, \ldots, x_d)$. Hence we can express each $\lambda_k(x)$ as

$$\lambda_k(x) = \sum_{i=1}^{d} w_{ki}x_i + w_{k0},$$

where w_{k0} is a nonnegative constant and w_{ki}'s are some real numbers for $i = 1, \ldots, d$. Define a $K \times d$ matrix W with entries w_{ki} and let S be the $d \times K$ matrix whose k-th column is the stoichiometry vector

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ζ_k for reaction k. Also let w_0 be the d-dimensional vector whose k-th component is w_{k0}. Regarding each vector as a column-vector, for any state-vector x we can write

$$\sum_{k=1}^{K} \lambda_k(x)\zeta_k = SWx + Sw_0,$$

which allows us to express the first equation in (2), with $Z_1(t) \equiv 0$, as

$$\frac{d\mathbb{E}[X(t)]}{dt} = SW\mathbb{E}[X(t)] + Sw_0. \tag{6}$$

This linear system of ODEs describes how the first-order moments of the open-loop network will evolve without the control action. If the matrix SW only has eigenvalues with negative real parts (we say in this case that the matrix SW is Hurwitz stable), then this system is *asymptotically stable* and the first-order moment vector $\mathbb{E}[X(t)]$ converges as $t \to \infty$. To fulfil our control objective, it is necessary that this system be asymptotically stable. This is because our controller can only act positively on the open-loop network and hence it cannot stabilize an unstable system.

Recall that $\mathcal{S} \subset \mathbb{N}_0^d$ is the state-space for the Markovian dynamics of the open-loop reaction network. This state-space is irreducible if any state in $\mathcal{S}$ can be reached from any other state in $\mathcal{S}$ by a sequence of reactions having positive propensities at all the intermediate states. A simple example is the single-species birth-death process for which from any state value we can reach any larger state value by a sequence of birth reactions and any smaller state value by a sequence of death reactions. The irreducibility of the state-space is a necessary condition for ergodicity [16], which holds for many reaction networks in the literature; see e.g. [10]. For a unimolecular open-loop network with an irreducible state-space $\mathcal{S}$, the asymptotic stability of the linear system (6) is equivalent to the ergodicity of the open-loop network.

However in order to ensure the ergodicity of the closed-loop network, we need a couple of other conditions that are both very natural for our control problem. The first condition is that the open-loop system be *output controllable* [20], which simply means, in our case, that the molecular count of the regulated species $\boldsymbol{X_\ell}$ responds to changes in the molecular count of the actuated species $\boldsymbol{X_1}$. Mathematically, we can express this condition as

$$[(SW)^{-1}]_{\ell 1} \neq 0 \tag{7}$$

where $[(SW)^{-1}]_{\ell 1}$ is the component at column 1 and row ℓ of the inverse of matrix SW. The second condition that we need is that the set-point $\mu^* = \mu/\theta$ should be accessible by the dynamics of $\mathbb{E}[X_\ell(t)]$. Since our controller can only act positively on the system (6), this accessibility condition can fail if some components of the input vector Sw_0 are *too large* (see also the discussion below Theorem **??** in the supplementary material). Technically we can check this accessibility condition by ensuring that there exists a positive constant c and a d-dimensional vector $v = (v_1, \ldots, v_d)$ with positive entries such that each component of the vector $v^T(SW + cI_d)$, where I_d is the $d \times d$ identity matrix, is strictly negative and

$$\mu^* = \frac{\mu}{\theta} > \frac{v^T S w_0}{c v_\ell}. \tag{8}$$

As discussed in [10], the above condition can be checked using linear programming techniques.

The following result, proved in Section **??** of the supplementary material, provides conditions under which a *unimolecular* open-loop reaction network can be controlled using the controller network (1):

Theorem 2 (Unimolecular Network Case). *Suppose that the open-loop reaction network* $(\boldsymbol{X}, \lambda, \eta)$ *is unimolecular and its state-space* $\mathcal{S}$ *is irreducible. Furthermore assume that the linear system* (6) *of*

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ODEs is asymptotically stable (SW is Hurwitz stable) and output controllable (condition (7) *holds), and that the desired set-point μ/θ is accessible (condition* (8) *holds).*

Then, for any $k, \eta > 0$, the closed-loop reaction network is ergodic and asymptotic set-point tracking is achieved, i.e. $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$. Moreover, the steady-state values of the first-order moments $\mathbb{E}[X(t)] = (\mathbb{E}[X_1(t)], \ldots, \mathbb{E}[X_d(t)])$ and $\mathbb{E}[Z_1(t)]$ are

$$\lim_{t\to\infty} \mathbb{E}[X(t)] = \frac{\mu(SW)^{-1}e_1}{\theta[(SW)^{-1}]_{\ell 1}} \quad \text{and} \quad \lim_{t\to\infty} \mathbb{E}[Z_1(t)] = \frac{-\mu}{\theta[(SW)^{-1}]_{\ell 1}}.$$

A more general version of this result and its extension to a class of bimolecular networks are provided in Section **??** and Section **??** of the supplementary material. In light of Theorem 2, several favorable properties for the controller and the closed-loop network can now be highlighted and expounded.

Ergodicity, set-point tracking and bounded first- and second-order moments. These are the main properties sought in our statement of control objectives. Moreover, as stated in Theorem 2, the average population-sizes of species $\boldsymbol{X_1}, \ldots, \boldsymbol{X_d}$ and $\boldsymbol{Z_1}$, at steady-state, are uniquely defined by the set-point μ/θ and the parameters of the open-loop network, implying that these quantities are also regulated by our antithetic integral controller.

Robustness. Robustness is a fundamental requirement which ensures that some properties for the closed-loop network are preserved, even in the presence of model uncertainties. This concept is critical in biology as the environment is fluctuating or noisy and only poorly known models are typically available. The obtained results can automatically guarantee the preservation of all the properties stated in Theorem 2, even in such constraining conditions.

Well-behaved single-cell tracking dynamics. Ergodicity ensures that the population average at stationarity is equal to the asymptotic value of the time-average of any single-cell trajectory; see e.g. [10]. We can therefore conclude that the proposed controller achieves two goals simultaneously, as it can ensure robust set-point tracking at both the population and single-cell levels. As a consequence, the controller will also ensure single-cell set-point tracking in the presence of cell events such as cell-division when certain conditions are met (see Section **??** of the supplementary material for more details).

Innocuousness of the controller. An inaccurate implementation of controller parameters may sometimes lead to an unstable behavior for the closed-loop system. However, the fact that the conditions of Theorem 2 are independent of k and η tells us that the proposed controller will both preserve the ergodicity of the (possibly poorly known) open-loop network and ensure set-point tracking/adaptation regardless of the values of its parameters, provided that the open-loop network satisfies the conditions of Theorem 2. This property is crucial in biology, as identifying models and implementing specific reaction rates (even approximately) are difficult tasks. This peculiar and non-standard property is referred here as *innocuousness* and it is illustrated in Figure 4, where the deterministic and stochastic dynamics of the same controlled networks are compared (see also Section **??** of the supplementary material).

Low metabolic load. Even if the controller works in the low copy-number regime, it does not necessarily imply a low metabolic load for the cell. Indeed if there is fast creation and annihilation of the controller species $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$, then it will result in many energy consuming futile cycles, which can impose a heavy metabolic burden on the cell, even though the dynamics is still in the low copy-number regime. However, it can be shown (see Section **??** of the supplementary material) that the power consumption at stationarity of the controller reactions, denoted by $\bar{P}$, can be expressed in the case of unimolecular networks as

$$\bar{P} = \mu\left(\alpha_1 + \alpha_2 + \alpha_3\right) + \frac{\mu}{\theta}\frac{\alpha_4}{|[(SW)^{-1}]_{\ell 1}|} \tag{9}$$

where $\alpha_1, \alpha_2, \alpha_3$ and α_4 are (positive) weights associated with the reference, measurement, comparison

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and actuation reactions, respectively, that represent the energy cost of each reaction. Interestingly, only the first three terms depend on μ while the last term depends on the ratio μ/θ and some network parameters. This last term, however, would also be present in the case of the production of $\boldsymbol{X_1}$ at a constitutive rate equal to $\mu/(\theta|e_\ell^T(SW)^{-1}e_1|)$ (which would lead to the same steady state value for the controlled output, but no adaptation properties). Hence, the effective metabolic load of our controller is equal to the first three terms and is only proportional to μ. A low metabolic load can therefore be easily achieved by first setting μ to a small value and then adjusting the set-point value with θ.

Circumventing moment closure difficulties. Finally, we emphasize that using the proposed approach, the moment closure problem does not arise, as the main conclusions (e.g. ergodicity, set-point tracking and robustness) directly follow from stochastic analysis tools and the structure of the controller, thereby avoiding altogether the framework of the moment equations (see Section **??** of the supplementary material).

Application to Gene Expression Control: set-point Tracking and Perfect Adaptation

The goal of this example is to demonstrate that set-point tracking and perfect adaptation can be ensured with respect to any change in the parameters of the gene expression network

$$\boldsymbol{X_1} \xrightarrow{\gamma_1} \emptyset, \quad \boldsymbol{X_1} \xrightarrow{k_2} \boldsymbol{X_1} + \boldsymbol{X_2}, \quad \boldsymbol{X_2} \xrightarrow{\gamma_2} \emptyset \tag{10}$$

where $\boldsymbol{X_1}$ denotes the mRNA and $\boldsymbol{X_2}$ the measured/regulated species (see Fig. 2). The following result is proved in Section **??** of the supplementary material:

Proposition 3. *For any positive values of the parameters $k, k_2, \gamma_1, \gamma_2, \eta, \theta$ and μ, the controlled gene expression network* (1)-(10) *is ergodic, has bounded and globally converging first- and second-order moments and*

$$\mathbb{E}[X_2(t)] \to \frac{\mu}{\theta} \quad \text{as} \quad t \to \infty. \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

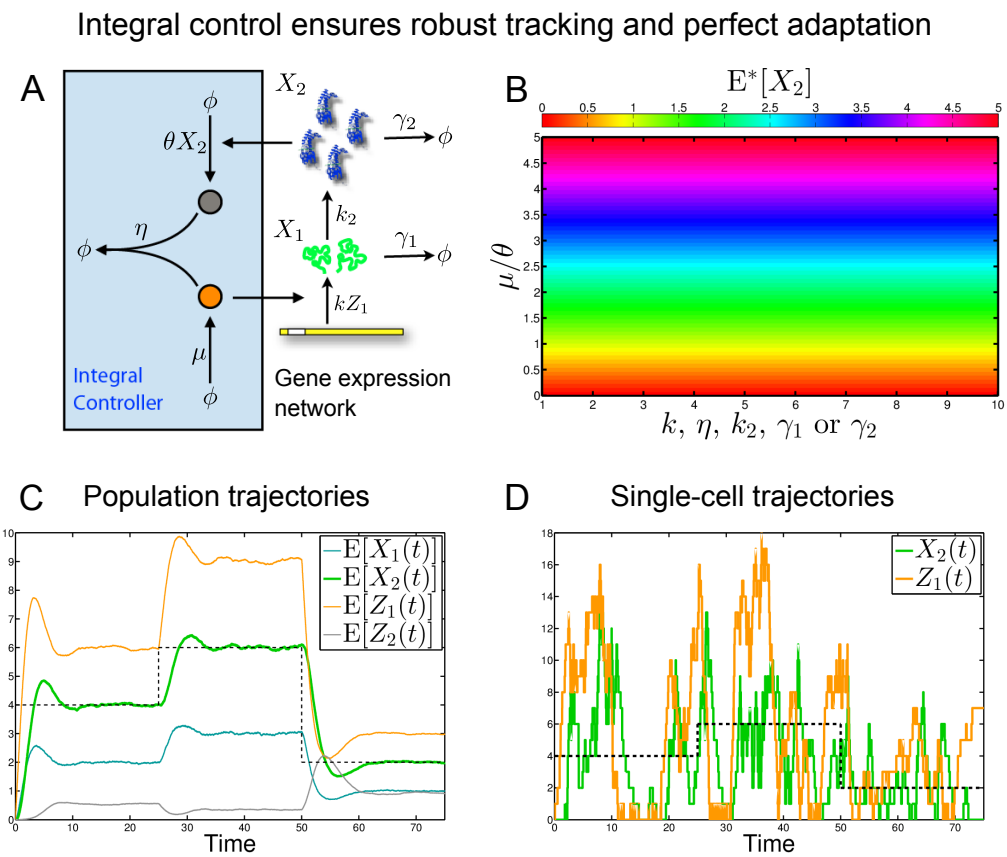


Figure 2: **A.** The controlled gene expression network (10) with the proposed antithetic integral controller (1). **B.** The closed-loop reaction network shows perfect adaptation (at stationarity) with respect to any changes in the parameters of the network as we have that $\mathbb{E}_\pi[X_2] = \mu/\theta$ for any values of the parameters k, η, k_2, γ_1 and γ_2 where $\mathbb{E}_\pi[X_2]$ denotes the mean number of molecules of $\boldsymbol{X_2}$ at stationarity. **C.** The controlled-output $\mathbb{E}[X_2(t)]$ of the closed-loop network tracks the reference value (in black-dash). The mean population of input species $\mathbb{E}[Z_1(t)]$ adapts automatically to changes in the reference value $\mu^* = \mu/\theta$ without requiring re-implementation. **D.** Single-cell trajectories, although strongly affected by noise, still have an underlying regularity ensuring the convergence of the moments at the population level. All simulations have been performed using Gillespie's stochastic simulation algorithm with the parameters $k = 1$, $\gamma_1 = 3$, $k_2 = 2$, $\gamma_2 = 1$, $\theta = 1$ and $\eta = 50$.

However, when we consider a Hill-type static control scheme of the form

$$\emptyset \xrightarrow{f(X_2)} \boldsymbol{X_1} \quad \text{with} \quad f(X_2) = \frac{\alpha K^n}{K^n + X_2^n} \tag{12}$$

where K, α are positive parameters and n is a positive integer, we obtain the results depicted in Fig. 3. We can see that, as opposed to the antithetic stochastic integral controller, perfect adaptation is not ensured by the Hill-type static controller. Note that even though the comparison is only made for the gene expression network and this specific choice for the Hill-type static controller, it is a matter of fact that, in general, such controllers can not ensure perfect adaptation; see Section **??** of the supplementary material for some theoretical arguments and different control schemes.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

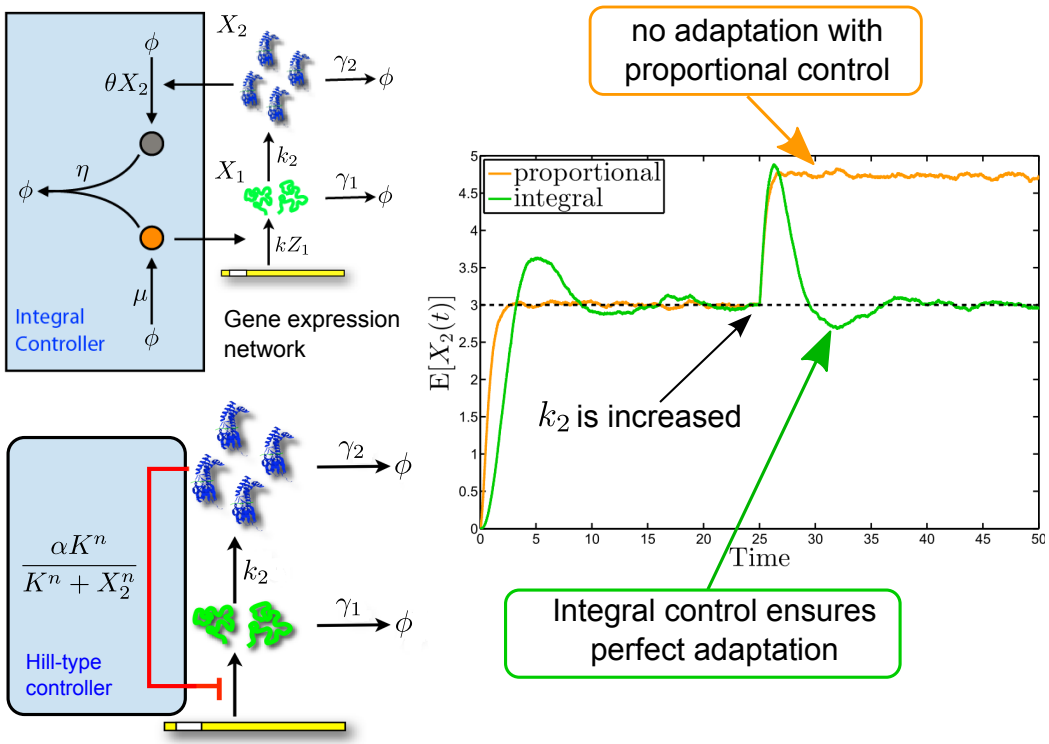


Figure 3: Comparison between the proposed antithetic integral controller (1) and the Hill-type static controller (12). The simulation is performed using parameters initialized to $\mu = 3$, $\theta = 1$, $k = 1$, $\gamma_1 = 3$, $k_2 = 3$, $\gamma_2 = 1$ and $\eta = 50$ for the antithetic integral controller (1) and $n = 1$, $\alpha = 8.22$ and $K = 3$ for the Hill-type static controller. The averaging is performed over 8000 cells simulated with Gillespie's stochastic simulation algorithm. At $t = 25$s, the value of k_2 jumps from 3 to 6. While the proposed antithetic integral controller shows perfect adaptation, the Hill-type static controller is unable to return to the mean value of the population of $\boldsymbol{X_2}$ before the stimulus. This demonstrates the advantage of the integral feedback strategy over the Hill-type strategy.

Noise as a Stabilizing Agent: Stochastic vs. Deterministic Population Control

Here we demonstrate a striking effect of noise as an agent for dynamic stabilization at the population level. We do this by comparing here the results that we obtain to those we would have obtained in the deterministic setting (see Fig. 4). To this aim, we again consider the gene expression network (10) and we set $k_2 = \gamma_1 = \gamma_2 = 1$ for simplicity. We then get the deterministic model and stochastic mean model depicted in Fig. 4-**A** and Fig. 4-**B**, respectively. It is important to emphasize that the deterministic model represents here the evolution of the mean concentration of the species over a population of deterministically behaving cells with identical initial concentrations. On the other hand, the stochastic mean model represents the evolution of the mean number of the species over a population of stochastically behaving cells with identical initial molecular counts. Note that, by virtue of the ergodicity property, the stochastic mean model will always converge to its unique steady-state value regardless of the different initial conditions for the individual cells. This property does not hold in general for a deterministic model representing the average concentrations of a population of deterministically oscillating cells. The stochastic mean model has been obtained using the identity $\mathbb{E}[Z_1 Z_2] = \mathbb{E}[Z_1]\mathbb{E}[Z_2] + \mathrm{Cov}(Z_1, Z_2)$ where

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

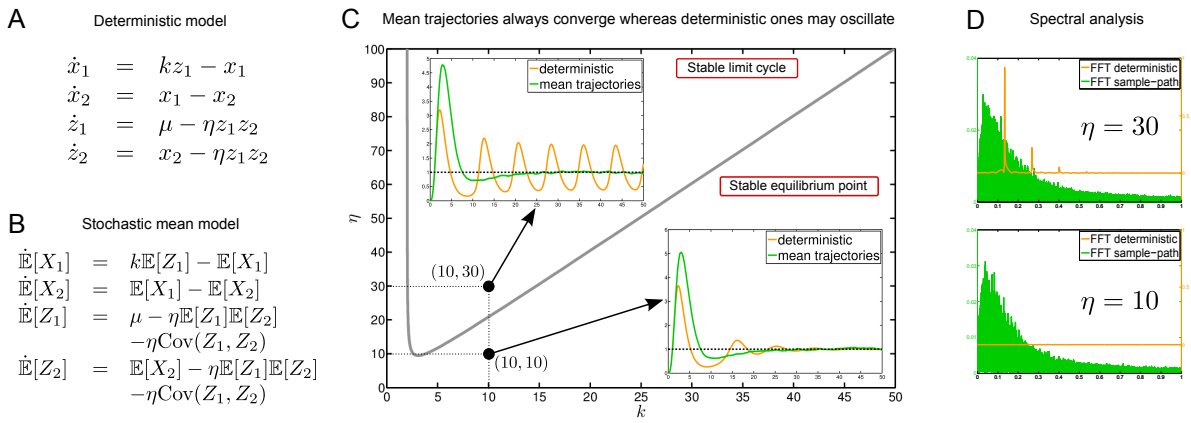


Figure 4: **A.** Deterministic model for the gene expression network (10) with $k_2 = \theta = \gamma_1 = \gamma_2 = 1$. **B.** Mean model for the gene expression network (10) with $k_2 = \theta = \gamma_1 = \gamma_2 = 1$. **C.** The deterministic dynamics bifurcates from a unique stable equilibrium point when the controller parameters (k, η) are chosen below the bifurcation curve into a stable limit-cycle when the controller parameters are chosen above. The first-order moments, however, always converge to the desired steady state value for the regulated species, here $\mu = 1$, regardless of the values of the controller parameters. This can be explained by the presence of the stabilizing covariance term in the model for the stochastic means. **D.** While the frequency content at stationarity of the deterministic dynamics dramatically changes when crossing the bifurcation curve, the frequency content of the sample-paths remains qualitatively the same. In this regard, the controller can be considered to perform the same way in both cases. This demonstrates the superiority and the central role of the noise in the stabilizing properties of the proposed stochastic integral controller.

the covariance term is nonzero as the random variables are not independent. If such a term would be zero, then we would recover the deterministic dynamics, but, due to noise, we can see in Fig. 4-**C** that while the deterministic dynamics may exhibit oscillations, the dynamics of the first-order moment is always globally converging to the desired steady-state value. As a final comment, we note that if we were closing the moment equations in Fig. 4-**C** by neglecting the second-order cumulant, then we would fail in predicting the correct behavior of the first-order moments. This demonstrates the central role of the noise in the stabilizing properties of the proposed stochastic antithetic integral controller. The noise can hence be viewed as here a stabilizing agent through the randomness it adds to the dynamics, allowing then for their systematic compensation at the population level, regardless of the initial conditions and the system parameters. This phenomenon is entirely due to stochasticity and it does not generically occur in the deterministic setting.

Discussion.

A general control theory for stochastic biochemical reaction networks with tailored mathematical concepts and tools has been missing. We believe that a well-grounded *biomolecular control theory* would enable a deeper understanding of biological regulation at the molecular level and could pave the way for an efficient and systematic rational design of synthetic genetic circuits that function to regulate and steer cellular dynamics at the molecular level. For such circuits, we propose the term "cybergenetic", which combines the genetic nature of the system with its cybernetic function of dynamic steering and control.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this article, we take a first step in the development of such a theory by addressing one of the central dynamic control motifs: integral feedback. The methods we developed are the product of a synthesis of ideas from control theory, probability theory, linear algebra and optimization theory. Even though our findings are specific to the class of integral controllers we consider, they may serve as the foundation on which more general *biomolecular control theory* can be developed – one that deals with a larger class of stochastic dynamic controllers and networks. Indeed, numerical experiments performed on more general networks lying outside the scope of the developed theory tend to support this claim (see Section **??** of the supplementary material).

Although the proposed control motif served as a frame around which the theoretical ideas of molecular stochastic control were developed, a cybergenetic circuit implementing such a motif may be of biological significance in its own right, a possibility which that we explore next for both endogenous and synthetic regulation. Indeed, the simplicity of the mechanism and the remarkable regulation properties that it confers raises the question whether such a motif could have evolved for the purpose of endogenous regulation. The species $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ may be RNA or protein, but they must act to effectively "annihilate" each other. Of course this annihilation need not be physical, as long as the two species act to render each other functionless. This could occur, for example, as a result of irreversible binding of $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$, where the new complex effectively sequesters both from performing their function. Remarkably, one finds just such a possibility in the sigma-factor-mediated *E. coli* regulatory system. One example is the sigma factor $\boldsymbol{\sigma^{70}}$, which binds to RNA polymerase core enzyme ($\mathbf{E}$) enabling it to recognize the promoter of a host of housekeeping genes (see Figure 5). It is well known that $\boldsymbol{\sigma^{70}}$ has a corresponding anti-sigma factor (Rsd) [26], denoted here by $\boldsymbol{\bar{\sigma}^{70}}$, which binds $\boldsymbol{\sigma^{70}}$ very tightly, sequestering it away from $\mathbf{E}$. Transcription of the Rsd gene is itself controlled by a $\boldsymbol{\sigma^{70}}$-dependent promoter [12]. This architecture fits very well our integral feedback regulation motif and strongly suggests that the core enzyme complex with $\boldsymbol{\sigma^{70}}$ ($\boldsymbol{E\sigma^{70}}$) is the object of tight regulation. This in turn points to a corresponding regulation of the expression of all the housekeeping genes whose promoters are recognized by $\boldsymbol{\sigma^{70}}$. When binding reaction rates in the literature [5, 15] are used for $\boldsymbol{\sigma^{70}}$ and RNAP along with estimated average total numbers, one comes up with a very low (< 1) average copy number of *free*-$\boldsymbol{\sigma^{70}}$ per cell, which would indicate that the system operates in the stochastic regime. Intriguingly, the anti-$\boldsymbol{\sigma^{70}}$ factor is known to be involved in the transition to stationary phase, during which time its concentrations are significantly elevated, leading to the down-regulation of the housekeeping genes. One way to achieve this is through the increase of the rate of transcription of anti-sigma factor Rsd during transition to stationary phase, leading to the decrease in the set-point for $\boldsymbol{E\sigma^{70}}$ (in our analysis this corresponds to the quantity μ/θ). Supporting this hypothesis is the fact that transcription of the $\boldsymbol{\sigma^{70}}$ gene is known to be controlled by a $\boldsymbol{\sigma^{38}}$-dependent promoter [12], where $\boldsymbol{\sigma^{38}}$ is recognized as the master regulator for adaptation to stationary phase transcription. While the full regulatory details of $\boldsymbol{\sigma^{70}}$ remain unknown, it is quite likely that $\boldsymbol{\sigma^{70}}$ and $\boldsymbol{\bar{\sigma}^{70}}$ interaction as described here plays a central part in that regulation. It is also quite possible that such a regulatory motif is not uncommon in biology, as several anti-sigma factors have been found in a number of bacteria, including *E. coli* and Salmonella, as well as in the T4 bacteriophage.

Beyond endogenous circuits, our cybergenetic motif presents opportunities for applications in synthetic biology. Until now most of the synthetic regulatory circuits have relied on proportional action–a control scheme that fails to ensure perfect adaptation in many practical situations. Moreover, existing theoretical studies of synthetic biological circuits mainly considered the deterministic setting, and hence they implicitly assumed large molecular abundances. However, the implementation of control circuits that rely on high component abundances severely impinges on the host circuit's material and energy resources, leading to increased metabolic burden which can affect both function and viability. Fortunately, this is largely avoidable, as effective control involves mostly information processing, which in principle requires little energy and material resource consumption. The novel regulatory motif that we propose exhibits characteristics that provably ensure robust stability, robust set-point tracking, and robust perfect adaptation for the controlled network and is achieved with a low metabolic cost and with molecular

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

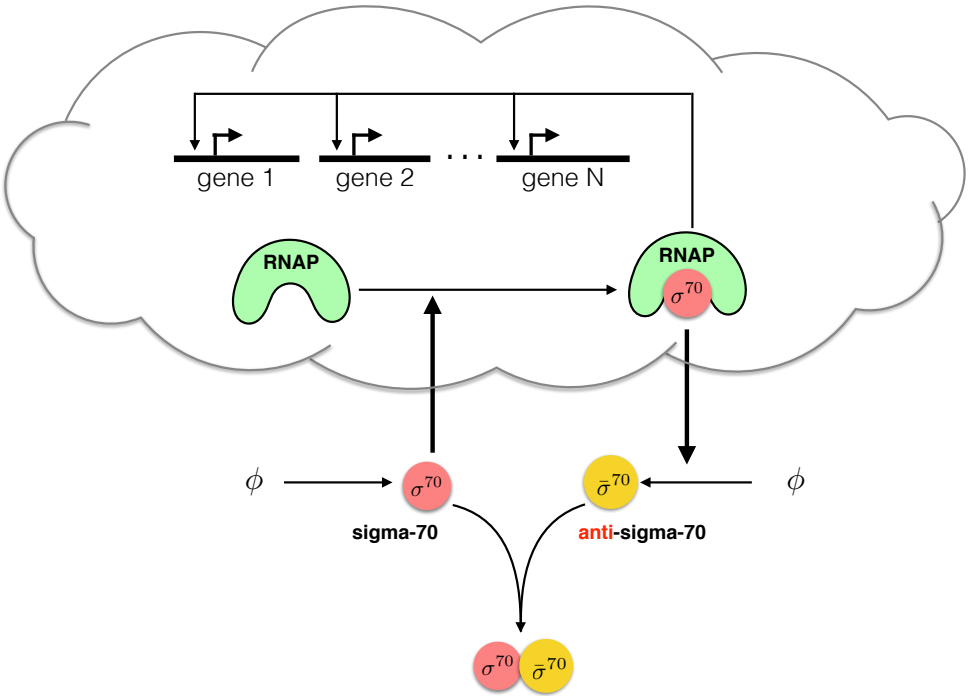


Figure 5: **Regulation of the housekeeping genes in *E. coli*.** An endogenous circuit that may employ the control strategy proposed in this article is the $\boldsymbol{\sigma^{70}}$ regulation circuit. The sigma-factor $\boldsymbol{\sigma^{70}}$ binds RNA polymerase core in a complex that controls the expression of housekeeping genes during exponential growth conditions. The anti-$\boldsymbol{\sigma^{70}}$ factor, Rsd, whose gene is controlled by $\boldsymbol{\sigma^{70}}$, binds with a very strong affinity to $\boldsymbol{\sigma^{70}}$, sequestering it away from RNA polymerase. According to the theory put forth in this article, both interactions result in the tight regulation of the concentration of the $\boldsymbol{E\sigma^{70}}$ complex through negative feedback. While the average abundance of RNA polymerase is about 4600 molecules per cell [5], the average abundance of the complex $\boldsymbol{E\sigma^{70}}$ (regulated species) is approximately 700 molecules per cell [12]. At the same time, the abundance of *free*-$\boldsymbol{\sigma^{70}}$ is computed to be fewer than one molecule per cell.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

species that can have very low abundances. It can be used for both single-cell set-point tracking (on average) and for population control. Thanks to the innocuousness of the controller, it does not need to be fine tuned, and can therefore be used in many practical situations, e.g. when the controlled network is very poorly known. In this regard, the proposed controller maintains clear implementability advantages over controllers requiring parameter tuning. This latter property emerges from the random nature of the reactions, as its deterministic counterpart leads to oscillating trajectories when the controller parameters are located in a certain instability region. In spite of this, this controller can still be used in a deterministic setting even though some of the properties, such the innocuousness property, are lost. With this in mind, the proposed controller may find several applications within synthetic biology. An immediate one is the optimization of drug or fuel production in bioreactors; see e.g. [7]. Currently simple control strategies, such as proportional feedback or constitutive production, are used in these applications. By utilizing slightly more complex controllers, such as the one proposed here, dramatic improvements in the production process can be expected, thanks to their enhanced robustness properties. Another important application example is the design of insulators; see e.g. [18]. It has indeed been shown that loading effects are often detrimental to modular design. Insulators are therefore needed in order to preserve function modularity. The proposed controller can be used as a buffering element in order to drive the output of a module to the input of another one. It can also be used as a constant signal generator that can be used to act on a network to be analyzed. The amplitude can be tuned by acting on the reference, which can be modified from outside the cell using light-induced techniques [17].

The proposed controller, however, may have some drawbacks, as it seems to introduce some additional variance to the controlled process. Even though this extra variance is not detrimental to the current control objectives, it may be a problem if the goal is to reduce the variance over a cell population. Whether the variance can be reduced via a more optimal choice of parameters or through some additional controller reactions, remains a question for further research. In the end, some "extra" variance due to the controller may be unavoidable, as fundamental limitations to variance reduction with feedback [13] are likely to hold for any molecular control circuit. Even so, controller noise should not detract from the tremendous promise of designing novel stochastic control circuits at the molecular level, where the dynamic properties, benefits, and limitations seem to be exquisitely different from those at the macroscopic scale.

Acknowledgments

The authors are grateful to Adam Arkin for pointing out the similarity of the tight sigma/anti-sigma factor binding reactions to our feedback controller reactions. The authors acknowledge funding support from the Human Frontier Science Program Grant RGP0061/2011 and the Swiss National Science Foundation grant 200021-157129.

Conflict of Interest

No conflicts of interest.

References

[1] K. J. Åström and T. Hägglund. *PID Controllers: Theory, Design, and Tuning.* Instrument Society of America, Research Triangle Park, North Carolina, USA, 1995.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[2] M. Acar, J. T. Mettetal, and A. van Oudenaarden. Stochastic switching as a survival strategy in uctuating environments. *Nature Genetics*, 40(4):471–475, 2008.

[3] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397:168–171, 1999.

[4] D. Anderson and T. G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In H. Koeppl, D. Densmore, G. Setti, and M. di Bernardo, editors, *Design and analysis of biomolecular circuits - Engineering Approaches to Systems and Synthetic Biology*, pages 3–42. Springer Science+Business Media, 2011.

[5] S. Bakshi, A. Siryaporn, M. Goulian, and J. C. Weisshaar. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology*, 85(1):21–38, 2012.

[6] Y.-J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Soloveichik, and G. Seelig. Programmable chemical controllers made from DNA. *Nature nanotechnology*, 8:755–762, 2013.

[7] M. J. Dunlop, J. D. Keasling, and A. Mukhopadhyay. A model for improving microbial biofuel production using a synthetic feedback loop. *Systems and Synthetic Biology*, 4(2):95–104, 2010.

[8] H. El-Samad, J. P. Goff, and M. Khammash. Calcium homeostasis and parturient hypocalcemia: An integral feedback perspective. *Journal of Theoretical Biology*, 214:17–29, 2002.

[9] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1997.

[10] A. Gupta, C. Briat, and M. Khammash. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLOS Computational Biology*, 10(6):e1003669, 2014.

[11] J. P. Hespanha. Moment closure for biochemical networks. In *3rd International Symposium on Communications, Control and Signal Processing*, pages 142–147, St. Julian's, Malta, 2008.

[12] M. Jishage and A. Ishihama. Transcriptional organization and in vivo role of the escherichia coli rsd gene, encoding the regulator of rna polymerase sigma d. *Journal of Bacteriology*, 181(12):3768–3776, 1999.

[13] I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467:174–178, 2010.

[14] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. Defining network topologies that can achieve biochemical adaptation. *Cell*, 138:760–773, 2009.

[15] H. Maeda, N. Fujita, and A. Ishihama. Competition among sevn Escherichia coli σ subunits: relative binding affinities to the core RNA polymerase. *Nuclear Acid Research*, 28(18):3497–3503, 2000.

[16] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Prob.* , 25:518–548, 1993.

[17] A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros. In silico feedback for in vivo regulation of a gene expression circuit. *Nature Biotechnology*, 29:1114–1116, 2011.

[18] D. Mishra, P. M. Rivera-Ortiz, A. Lin, D. Del Vecchio, and R. Weiss. A load driver device for engineering modularity in biological networks (in press). *Nature Biotechnology*, 32(12):1268–1275, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted November 23, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[19] D. Muzzey, C. A. Goméz-Uribe, J. T. Mettetal, and A. van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 97(9):160–171, 2009.

[20] K. Ogata. *Modern control engineering*. Prentice Hall, Upper Saddle River, 1970.

[21] K. Oishi and E. Klavins. Biomolecular implementation of linear I/O systems. *IET Systems Biology*, 5(4):252–260, 2010.

[22] J. Paulsson, O. G. Berg, and M. Ehrenberg. Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7148–7153, 2000.

[23] G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327:1389–1391, 2010.

[24] J. A. Stapleton, K. Endo, Y. Fujita, K. Hayashi, M. Takinoue, H. Saito, and T. Inoue. Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synthetic Biology*, 1(3):83–88, 2012.

[25] T. Tian and K. Burrage. Stochastic models for regulatory networks of the genetic toggle switch. *Proc. Natl. Acad. Sci.*, 103(22):8372–8377, 2006.

[26] L. G. Treviño-Quintanilla, J. A. Freyre-Gonzáles, and I. Martinez-Flores. Anti-sigma factors in e. coli: Common regulatory mechanisms controlling sigma factors availability. *Current Genomics*, 14:378–387, 2013.

[27] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences of the United States of America*, 109(35):14271–14276, 2012.

[28] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillator. *Proc. Natl. Acad. Sci.*, 99(9):5988–5992, 2002.

[29] 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 of the United States of America*, 97(9):4649–4653, 2000.