bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

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

June 11, 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, synthetic homeostatic circuits have received far less attention. The tight regulation of the abundance of cellular products and intermediates in the noisy environment of the cell is now recognised as a critical requirement for several biotechnology and therapeutic applications. 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 synthetic homeostatic circuits. Using these ideas, we propose a new regulation motif that implements an integral feedback strategy which can generically and effectively regulate a wide class of reaction networks. By combining tools from probability and control theory, we show that the proposed control 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 without any prior knowledge of reaction rates. Moreover, our proposed control 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. Several examples demonstrate the potential of the approach.

Perfect adaptation is that property of a biological system (e.g. a cell), that 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 [25] 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 [7] 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 [16], 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. Adaptation, however, may not be necessarily related to integral control as some theoretical studies have suggested [12, 20].

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

*Corresponding author; e-mail: mustafa.khammash@bsse.ethz.ch

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

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 [6]. Another synthetic negative feedback loop is also designed in [21] 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 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 tracking and robust perfect adaptation. Unlike the deterministic setting, however, tracking and adaptation robustness must be maintained not only with respect to model parameters, but also for the highly fluctuating species abundances. One way to attempt the construction of a "stochastic integral feedback" is to use statistical moments 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 [14]. 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. [10]. Similar difficulties arise if one works with the chemical master equation; see e.g. [8].

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

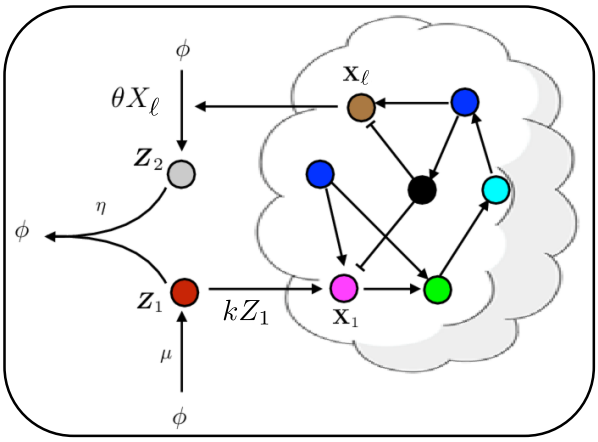

Figure 1: Schematic representation of the closed-loop network controlled by the proposed integral controller (1). The controller (left side) acts on the network (right side) by influencing the rate of production of the actuated species $\boldsymbol{X_1}$ by means of the control input species $\boldsymbol{Z_1}$. 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 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 integral feedback network". 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 two of the controller parameters, and is independent of the network parameters. In this respect, the closed-loop network exhibits stability, robust 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 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 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. This provides a clear example where the intrinsic stochastic noise is very 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 [19], noise-induced oscillators [24] and noise-induced switches [2, 22], have appeared in the literature in recent years.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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 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 elegantly 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 theoretical results are finally demonstrated by numerical simulations.

Results

The open-loop network.

We describe here the reaction network we aim to control. Let us consider the Markovian model of a reaction network with mass-action kinetics involving d molecular species denoted by $\boldsymbol{X_1}, \ldots, \boldsymbol{X_d}$. Under the well-stirred assumption [4], the state of the system is given, at any time, by the vector of molecular counts of the d species. The state evolution is described 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* $\zeta_k \in \mathbb{Z}^d$. Here λ_k is called the *propensity function* of the k-th reaction and is assumed to verify that if, for any $x \in \mathbb{N}_0^d$, we have $x + \zeta_k \notin \mathbb{N}_0^d$, then $\lambda_k(x) = 0$. This ensures that molecular counts of all the species remain nonnegative throughout the dynamics.

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 ($\lambda_k(x) > 0$). 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 $\mathcal{S}$ [9]. Let $\{X(t) = (X_1(t), \ldots, X_d(t)) : t \geq 0\}$ be the continuous time Markov process representing the reaction dynamics with an initial state $x_0 \in \mathcal{S}$.

Actuated and regulated species.

From a control theoretic point of view, it is necessary to define input and output nodes of the above network. We assume here that the species $\boldsymbol{X_1}$ is the *actuated species* which is the species we can act on. The *regulated species* is $\boldsymbol{X_\ell}$, for $\ell \in \{1, \ldots, d\}$, is the species we would like to control. The way we act on the actuated species as well as the way we want to control the regulated species is explained in the next sections.

The control problem.

It is important to state now the control problem we are interested in.

Problem 1 *Find a controller such that, by suitably acting on the actuated species $\boldsymbol{X_1}$, we have the following properties for the closed-loop network (defined here as the interconnection of the network $(\boldsymbol{X}, \lambda, \zeta)$ defined above with the controller):*

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 to their unique stationary value;*
3. *we have that $\mathbb{E}[X_\ell(t)] \to \mu^*$ globally 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

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

network is well-behaved (in the sense that it reaches stationarity). The second requirement is more specific to stochastic processes, as even if the means converge, the variance can still go unbounded which would mean that the actual dynamics of the process (its sample-paths) is not well-behaved and, therefore, that the controller is of little practical utility. Finally, the third statement encapsulates our desired objective of perfect adaptation (or tracking), i.e. that the population mean approaches a fixed homeostatic value.

Controller network.

We propose the following controller network (see Fig. 1) inspired from the deterministic networks proposed in [18]:

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

Above $\boldsymbol{X_\ell}$ is the *measured species* which turn out to be 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*. More specifically, $\boldsymbol{Z_1}$ is the *reference species* since its birth rate depends on the set-point while $\boldsymbol{Z_2}$ is the *sensing species* since its production rate is proportional to the population of the measured species. The species $\boldsymbol{Z_1}$ also plays the role of the *control input species* since it acts on the birth rate of the *actuated species* $\boldsymbol{X_1}$. Note that the topology of the controller network belongs to a family of four control topologies (see Fig. S1 to Fig. S4 in the Supplementary material) depending on what is the control input species and what is the sensing species. In this regard, $\boldsymbol{Z_1}$ may not necessarily be the control input species. The proposed theory is, however, only valid for the above controller network and the analysis of the three other ones is left for future research.

Although inspired from [18], the above network has a different philosophy. Besides the fact that the current setting is stochastic, the main difference lies in the way how the network interacts with the environment. While the goal of [18] was the biomolecular implementation of linear input-output systems, the goal here is the control a chemical 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 [18].

We now clarify the role and meaning of each of these 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 of $\boldsymbol{X_\ell}$.

3. The third reaction implements the *comparison reaction* decreasing by one the respective populations of $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$, at a rate η that can be tuned. The main role of 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 *comparison and substraction operation* since when both $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ have positive populations (comparison), then we decrement their respective population, thereby preserving the same difference level $Z_1 - Z_2$.

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.

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

[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 August 17, 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.

The "hidden" integral action.

It seems important to identify the source of the integral action. From the stationary moments equations of the controller network (1)

$$\begin{array}{rcl} 0 & = & \mu - \eta \mathbb{E}^*[Z_1 Z_2] \\ 0 & = & \theta \mathbb{E}[X_\ell] - \eta \mathbb{E}^*[Z_1 Z_2] \end{array} \tag{2}$$

we get that $\mu - \theta \mathbb{E}^*[X_\ell] = 0$, and thus that $\mathbb{E}^*[X_\ell] = \mu/\theta$, where $\mathbb{E}^*$ denotes expectation at stationarity. Therefore, the controller automatically imposes the value μ/θ to $\mathbb{E}^*[X_\ell]$ regardless of the values of all the other parameters, which is the main rationale for integral control. In this regard, proving that the closed-loop network reaches stationarity will automatically imply that $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$, without the need for solving any moments equations.

Implementation of the controller.

The proposed controller (1) 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 very strong robustness properties which make its implementation much easier than other types of controllers which require the fine tuning of their reaction rates (see also Section S5.4 of the supplementary material). In-vitro control is also possible using, for instance, DNA strand displacement [5]. In-silico control [14, 23], finally, can also be considered whenever the population of regulated species $\boldsymbol{X_\ell}$ can be measured from the outside of the cell(s) using, for instance, time-lapse microscopy.

Unimolecular networks case.

The following result, proved in Section S6 of the supplementary material, establishes conditions under which a stochastic unimolecular reaction network can be controlled using the controller network (1):

Theorem 2 *Let us first define the matrices $S \in \mathbb{Z}^{d \times K}$ and $W \in \mathbb{R}^{K \times d}$, and the vector $w_0 \in \mathbb{R}^K_{\geq 0}$ as $S := \begin{bmatrix} \zeta_1 & \dots & \zeta_K \end{bmatrix}$ and $\lambda(x) =: Wx + w_0$. Define also e_i to be the d-dimensional vector with the i-th entry equal to 1 and 0 elsewhere.*

Assume now that the state-space of the reaction network is irreducible and that the positive linear system describing the dynamics of the first-order moments given by

$$\begin{array}{rcl} \dfrac{\mathrm{d}\mathbb{E}[X(t)]}{\mathrm{d}t} & = & SW\mathbb{E}[X(t)] + e_1 u(t) + Sw_0 \\ y(t) & = & e_\ell^T \mathbb{E}[X(t)] \end{array} \tag{3}$$

is asymptotically stable and has non-zero impulse-response $h(t) := e_\ell^T e^{SWt} e_1$, $t \geq 0$. Then,

1. *the closed-loop reaction network is ergodic;*
2. *the first- and second-order moments of $X(t)$ exist and are uniformly bounded and globally converging;*
3. *we have that*

   $$\mathbb{E}[X_\ell(t)] \to \mu/\theta \text{ as } t \to \infty$$

   provided that the condition

   $$\frac{\mu}{\theta} > \frac{v^T S w_0}{c v^T e_\ell}$$

   is satisfied for some scalar $c > 0$ and some vector $v \in \mathbb{R}^d_{>0}$ satisfying $v^T(SW + cI) \leq 0$.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

The assumptions of the above result can be interpreted in the following way. The irreducibility assumption on the state-space of the reaction network is a necessary condition for ergodicity [13] which holds for many reaction networks of the literature; see e.g. [9]. The assumption on the asymptotic stability is also necessary as integral control does not have any stabilizing effect, except in some very particular cases. Finally, the assumption that the impulse response h must not be identically zero is equivalent to saying that the system is *output controllable* [17] or, in the present case, that the system simply responds to the inputs that are applied to it. These conditions are therefore very natural for the current problem. More general versions of this result and its extension to a class of bimolecular networks are provided in Section S6 and Section S7 of the supplementary material.

Properties of the closed-loop network.

In light of Theorem 2, several striking properties for the closed-loop network and the controller itself can be stated.

Ergodicity, tracking and bounded first- and second-order moments. These are the main properties stated in the considered control problem, i.e. Problem 1.

Robustness. Robustness is another fundamental property ensuring that some properties for the closed-loop network are preserved, even in presence of model uncertainties. This concept is critical in biology as the environment is fluctuating (noise) and poorly known models are only available. The obtained results can automatically guarantee the preservation of all the properties stated in Theorem 2, even in such constraining conditions.

Single-cell behavior and population behavior. 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. [9]. We can therefore conclude that the proposed controller achieves two goals simultaneously as it can, at the same time, ensures robust tracking at both a population and a single-cell level. As a consequence, the controller will also ensure single-cell tracking in presence of cell events such as cell-division and cell-growth (see Section S8.5 of the supplementary material).

Innocuousness of the controller. *Innocuousness* is a non-standard property of the proposed controller meaning that it can safely be implemented to achieve the control objectives regardless of the values of its parameters k, η (the conditions of Theorem 2 are independent of k, η). This property is quite uncommon (see Section S5.4 of the supplementary material) as an incorrect implementation of control laws usually result in the incapacity of ensuring the control objectives. This is known as *fragility*. This suggests that the proposed controller can be implemented without any clear knowledge of both the network and the controller parameters. This is crucial in biology as identifying models and implementing specific reaction rates (even approximately) both remain an elusive task.

Circumventing moment closure difficulties. Finally, we emphasize that using the proposed approach, the moment closure problem does not arise as the conclusion (i.e. ergodicity, tracking and robustness) directly follows from stochastic analysis tools and the structure of the controller, thereby avoiding altogether the framework of the moment equations (see Section S8.6 of the supplementary material).

Proportional action vs. integral control.

To illustrate the difference between proportional and integral action, we consider here the following gene-expression network

$$\boldsymbol{X_1} \xrightarrow{\gamma_1} \emptyset, \quad \boldsymbol{X_1} \xrightarrow{k_2} \boldsymbol{X_1} + \boldsymbol{X_2}, \quad \boldsymbol{X_2} \xrightarrow{\gamma_2} \emptyset \tag{4}$$

where $\boldsymbol{X_1}$ denotes the mRNA and $\boldsymbol{X_2}$ the corresponding protein. We compare now by simulation (see Fig. 2) the performance of these controllers:

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

- The integral controller (1) where $\boldsymbol{X_1}$ is the actuated species and $\boldsymbol{X_2}$ the measured/regulated species.
- The proportional controller described by the reaction

$$\emptyset \xrightarrow{f(X_2)} \boldsymbol{X_1} \quad \text{with} \quad f(X_2) = \frac{\alpha K^n}{K^n + X_2^n} \tag{5}$$

  where K, α are positive parameters and n is a positive integer.

Even though the comparison is based on these specific networks, it is a matter of fact that proportional controller can not ensure adaptation; see Section S8.1 of the supplementary material for some theoretical arguments and different proportional schemes.

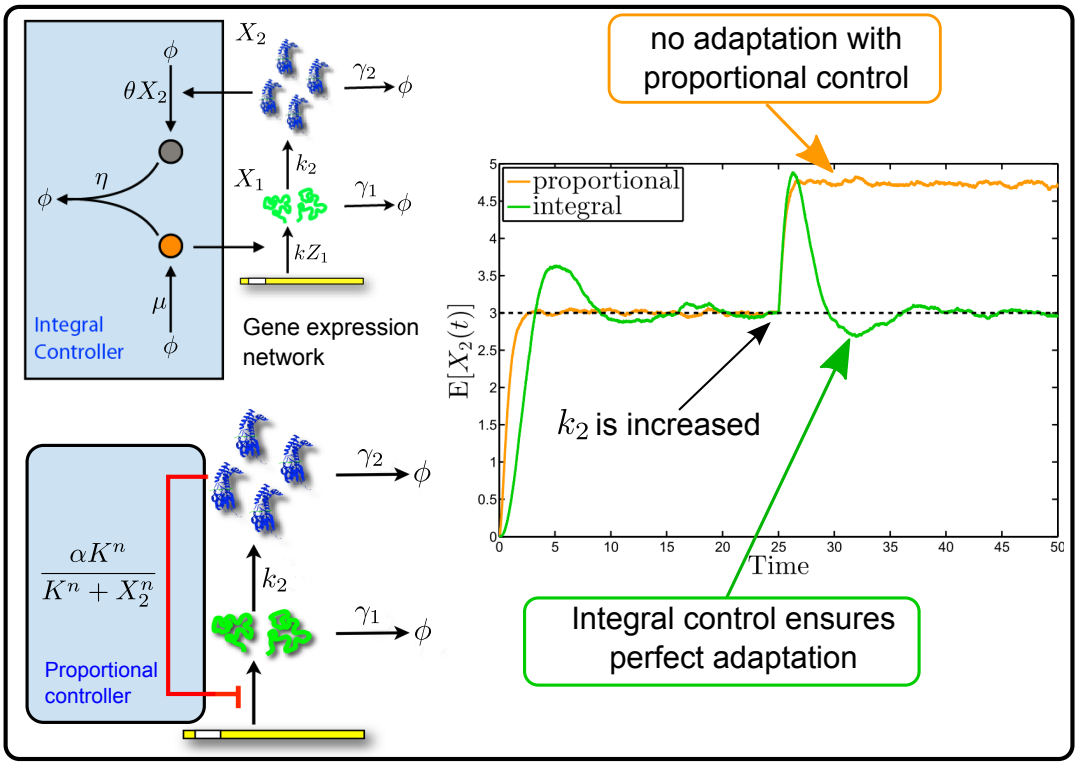

Figure 2: Comparison between the proposed integral controller (1) and the proportional feedback (5). The simulation is performed using parameters initialized to $\mu = 3$, $\theta = 1$, $k = 1$, $\gamma_1 = 3$, $k_2 = 3$, $\gamma_2 = 1$ and $\eta = 50$ for the integral controller (1) and $n = 1$, $\alpha = 8.22$ and $K = 3$ for the proportional 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 integral controller shows perfect adaptation, the proportional 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 proportional strategy.

Gene expression control - Output tracking and perfect adaptation.

The goal of this example is to demonstrate that tracking and perfect adaptation can be ensured with respect to any change in the network parameters for the gene expression network (4) where $\boldsymbol{X_1}$ is again

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

the actuated species and $\boldsymbol{X_2}$ the measured/regulated species (see Fig. 3). The following result is proved in Section S8.2 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* (4)-(1) *is ergodic, has bounded and globally converging first- and second-order moments and*

$$\mathbb{E}[X_2(t)] \to \frac{\mu}{\theta} \quad \text{as} \quad t \to \infty. \tag{6}$$

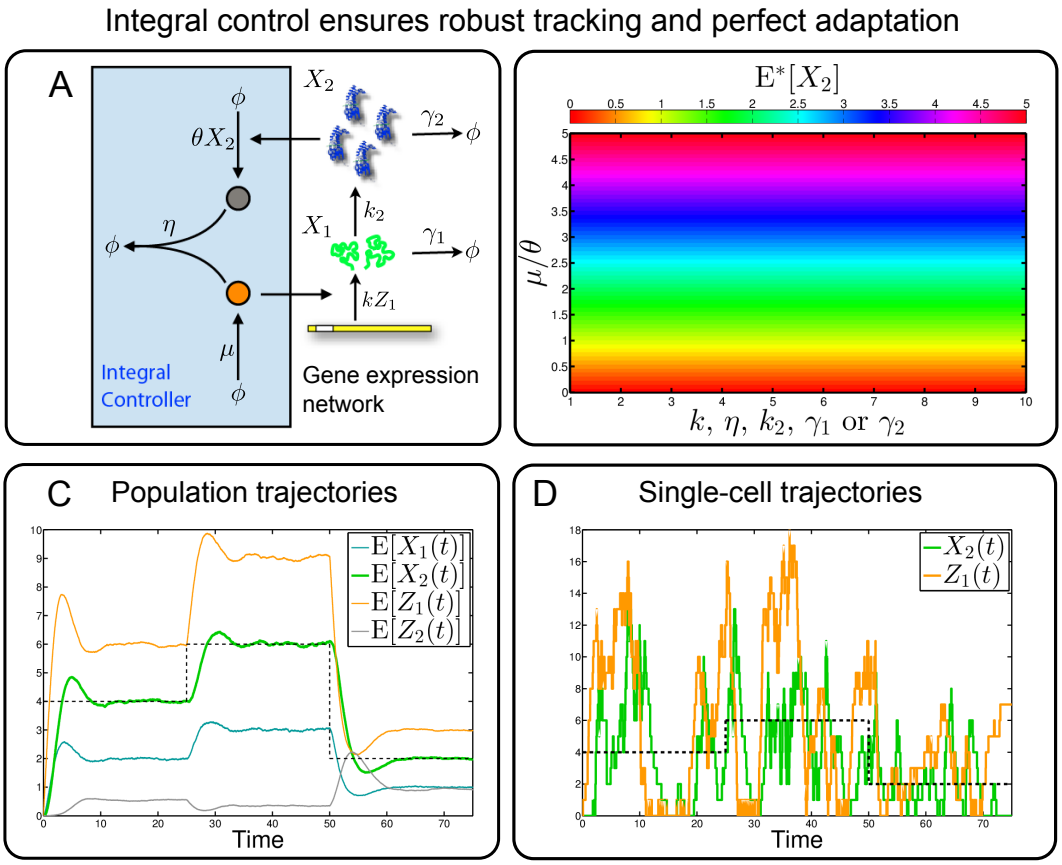

Figure 3: **A.** The controlled gene expression network (4) with the proposed integral controller (1). **B.** The closed-loop reaction network shows perfect adaptation (at stationarity) with respect to any changes in the parameters of the network as we have that $\mathbb{E}^*[X_2] = \mu/\theta$ for any values of the parameters k, η, k_2, γ_1 and γ_2 where $\mathbb{E}^*[X_2]$ 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 $\eta = 50$.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

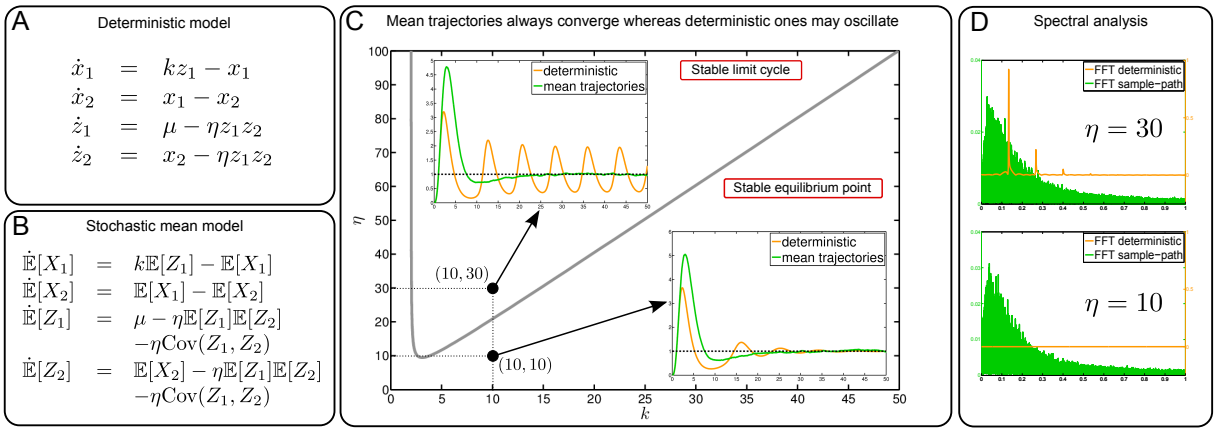

Figure 4: **A.** Deterministic model for the gene expression network (4) with $k_2 = \theta = \gamma_1 = \gamma_2 = 1$. **B.** Mean model for the gene expression network (4) with $k_2 = \theta = \gamma_1 = \gamma_2 = 1$. **C.** The deterministic dynamics bifurcates from a unique stable equilibrium point 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.

Deterministic vs. stochastic control.

It seems important to compare the results that we obtain here to those we would have obtained in the deterministic setting (see Fig. 4). To this aim, we consider again the gene expression network (4) to which we set $k_2 = \gamma_1 = \gamma_2 = 1$ for simplicity. We then get the deterministic and stochastic models depicted in Fig. 4-**A** and Fig. 4-**B**, respectively. 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 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 moments equation 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 integral controller.

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* could pave the way for an efficient and systematic rational design of synthetic in vivo regulatory motifs, in the same way that classical control theory opened the way for numerous novel applications in various engineering disciplines. In this article, we aimed to lay the foundation for such a theory by addressing one of the

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

central feedback 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 to develop a more general *biomolecular control theory* – 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 S8.7 of the supplementary material).

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 species 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 is concerned more about information processing which need not require high energy or material resources. The novel regulatory motif that we propose exhibit characteristics that provably ensure robust stability, robust tracking, and robust perfect adaptation for the controlled network, achieved with molecular species that can have very low abundances. It can be used for both single-cell tracking (in 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.

The proposed controller structure may find several applications. An immediate one is the optimization of drug or fuel production in bioreactors; see e.g. [6]. Currently naive control strategies, such as proportional feedback or constitutive production, are used in these applications. By utilizing more sophisticated 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. [15]. 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. Finally, the controller 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 [14]. One major benefit of the proposed controller, in this case, lies in its versatility as it does not need to be specifically designed for any reference value.

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 one's goal is to reduce the variance over a cell population. This, however, may be unavoidable, as fundamental limitations to variance reduction with feedback have been established [11]. When variance has to be reduced, cell-to-cell communication via quorum sensing might be a viable solution to compensate for the additional randomness the controller is introducing.

Conflict of Interest

No conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

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.

[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] 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.

[6] 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.

[7] 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.

[8] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

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

[19] 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.

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

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

Figures legends

Legend Figure 1. Schematic representation of the closed-loop network controlled by the proposed integral controller (1). The controller (left side) acts on the network (right side) 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.

Legend Figure 2. Comparison between the proposed integral controller (1) and the proportional feedback (5). 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 integral controller (1) and $n = 1$, $\alpha = 8.22$ and $K = 3$ for the proportional 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 integral controller shows perfect adaptation, the proportional 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 proportional strategy.

Legend Figure 3. A. The controlled gene expression network (4) with the proposed 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 $\mathbb{E}^*[X_2] = \mu/\theta$ for any values of

bioRxiv preprint doi: https://doi.org/10.1101/024919; this version posted August 17, 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.

the parameters k, η, k_2, γ_1 and γ_2 where $\mathbb{E}^*[X_2]$ 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 $\eta = 50$.

Legend Figure 4. A. Deterministic model for the gene expression network (4) with $k_2 = \theta = \gamma_1 = \gamma_2 = 1$. **B.** Mean model for the gene expression network (4) 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.