Control of protein concentrations in heterogeneous cell populations

A. Vignoni[†], D. A. Oyarzún[‡], J. Picó[†], and G.-B. Stan[‡]

***Abstract*— In this work we propose a synthetic gene circuit for controlling the variability in protein concentration at a population level. The circuit, based on the use of an intracellular nonlinear controller coupled to a cell-to-cell communication mechanism, allows for independent control of the mean and variance of a signalling molecule across cell population. Via a piecewise affine approximation of the nonlinearity, we provide set invariance results that imply the stability of the closed loop system. We also obtain closed-form expressions for the mean and variance as a function of the tuneable parameters of the controller. The predictions offered by the theoretical analysis are in agreement with numerical simulations performed with physiologically realistic parameters in *Escherichia coli*.**

I. Introduction

Since the seminal works in [1], [2], a number of biomolecular devices have been developed to perform circuit-like functions in living cells, including switches, pulse generators and logic gates [3]. Substantial efforts are being undertaken to scale up synthetic biology from individual modules to whole systems capable of executing complex functions [4].

An area of particular relevance is the design of collective cell behavior, whereby a prescribed population response results from the interaction between individual cells. A common approach to induce collective behaviors is to use cell-to-cell communication mechanisms. These typically rely on the quorum sensing machinery from *V. fischeri* and have been used for diverse purposes such as population synchronization [5], cell density control [6], engineered pattern formation [7] and the design of synthetic ecosystems [8].

Gene expression is an inherently stochastic process, and it is widely acknowledged that genetic noise plays a key role in cellular dynamics [9]. At a population level, the effect of noise becomes apparent by the fact that genetically identical cells produce the same protein at different concentrations. The variability in protein concentrations can be quantified with high-throughput technologies such as flow cytometry, which allow to characterize the variability in terms of the population histograms for the protein abundance [10].

Research in this area is partially supported by Spanish government (FEDER-CICYT DPI2011-28112-C04-01) and European Union (FEDER). Alejandro Vignoni thanks grants FPI/2009-21 and PAID-00-11/2714 of Universitat Politècnica de València. Dr Guy-Bart Stan and Dr Diego Oyarzún gratefully acknowledge the support of the EPSRC Centre for Synthetic Biology and Innovation (project EP/G036004/1) at Imperial College London.

[†]Institut d'Automàtica i Informàtica Industrial, Universitat Politècnica de València, Camí de Vera s/n, 46022, Valencia, Spain.

[‡]Centre for Synthetic Biology and Innovation and Department of Bioengineering, Imperial College London, SW7 2AZ, United Kingdom.

We thank Evan Olson and Dr Jeff Tabor from Bioengineering Department at Rice University and Gabriel Bosque from AI2 at Universitat Politècnica de València for fruitful discussions on the biology of this problem. We also thank reviewers whose comments helped to improve this work.

In this work we combine an intracellular feedback controller with a cell-to-cell communication mechanism designed to control the mean and variance of the signalling molecule Acyl-Homoserine Lactone (AHL) across a population of cells. AHL is an autoinducer molecule that diffuses in the extracellular medium and acts as a communication signal between cells. The feedback controller regulates the production of the protein LuxI, which in turn controls the synthesis of AHL (Fig. 1).

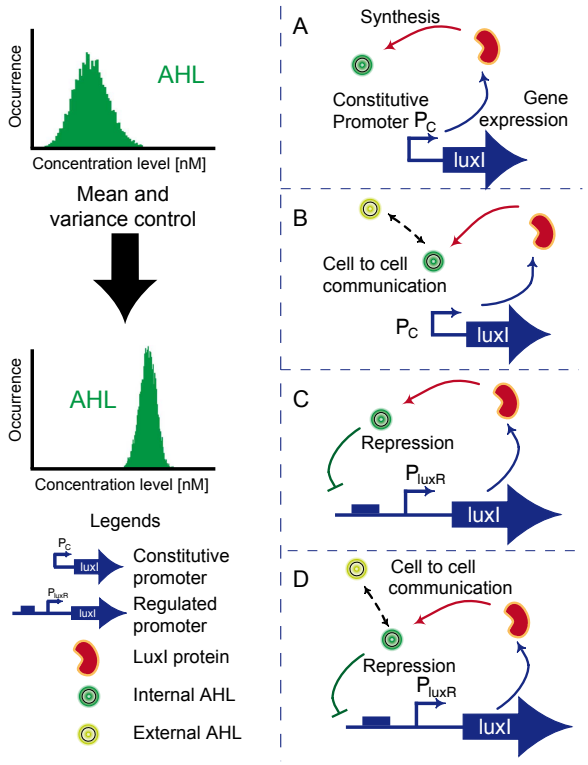

Fig. 1. Different architectures for controlling the AHL distribution across a population. The core principle is to use the PluxR promoter as an interface to control the production of LuxI (thus modulating AHL and the cell-to-cell communication). (**A**) Open loop production of LuxI without communication. (**B**) Open loop production of LuxI with communication. (**C**) Feedback-regulated production of LuxI without communication. (**D**) Feedback-regulated production of LuxI with communication.

This mentioned heterogeneity in a population of cells is usually modelled via deterministic ODEs with parameters sampled from a given probability distribution, which is sometimes termed as *extrinsic noise* [11]. The same approach was used in this work to account for variability across the population.

We first consider an ODE model for the intracellular

genetic circuit coupled with the dynamics of AHL export and uptake (Section II). The saturable behavior of promoter activity translates into a sigmoidal nonlinearity in the feedback controller. By approximating the nonlinearity with a piecewise affine function (Section III), we find conditions under which the system operates in a linear regime (Section IV). We rely on set-invariance results similar to those developed in [12] for continuous implementations of sliding mode control, and in [13] for a sliding mode reference conditioning scheme for coordination of multi-agents. Our main results are closed-form expressions for the mean and variance of AHL across the population (Section V). These indicate how a target mean and variance can be achieved independently by fine-tuning the controller parameters. The predictions offered by our theoretical analysis are in agreement with numerical simulations (Section VI) performed with physiologically realistic parameters of *Escherichia coli*.

II. System description

A. Cell-to-cell communication and feedback controller

The proposed circuit combines two engineered gene networks previously implemented in *E. coli*: a cell-to-cell communication system [14], and a synthetic repressible promoter [15], see Fig. 2. The cell-to-cell communication circuit uses components taken from the quorum sensing system of *V. fischeri* [16], [17]. The feedback circuit comprises a luxI gene under the control of the PluxR promoter. The protein LuxI is the AHL synthase. AHL in turn can bind the protein LuxR and form a complex that binds to the PluxR promoter and represses the expression of the luxI gene. The circuit therefore a negative feedback loop between the concentration of intracellular AHL and the expression of luxI gene.

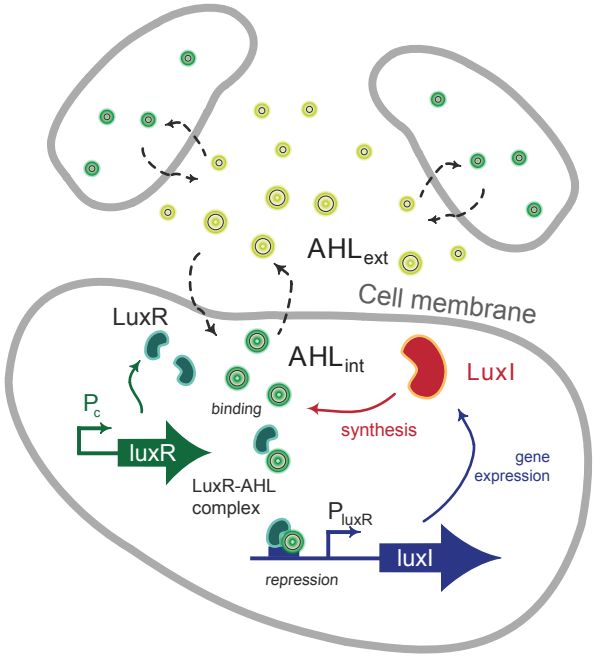

Fig. 2. Schematic of the intracellular feedback control circuit in one cell and the cell-to-cell communication mechanism.

B. Mathematical model

In a population of N cells, the concentrations of LuxI, internal AHL and external AHL can be modelled by the following set of ODEs

$$\dot{x}_1^i = \kappa_0 + \kappa_1 \frac{\theta^n}{\theta^n + {x_2^i}^n} - \gamma_1 x_1^i, \quad (1)$$

$$\dot{x}_2^i = \kappa_2 x_1^i - d\left(x_2^i - x_e\right) - \gamma_2 x_2^i, \quad (2)$$

$$\dot{x}_e = \frac{d_e}{N}\sum_{i=1}^{N}\left(x_2^i - x_e\right) - \gamma_e x_e, \quad (3)$$

where x_1^i is the concentrations of LuxI protein in the i^{th} cell, x_2^i is the concentration of internal AHL in the i^{th} cell, and x_e is the concentration of external AHL.

In equation (1) modelling the LuxI concentration, κ_0 is the tightness or basal expression of the promoter, and κ_1 is the dynamic range of the promoter. The regulatory effect of the promoter is modelled as a Hill-like function, whereby n is the Hill coefficient and θ is the half concentration constant or repression threshold.

In equation (2) modelling the internal AHL concentration, the kinetic constant κ_2 models first order AHL synthesis, whereas d is the internal transport constant. Both molecules, LuxI and internal AHL, are subject to first order degradation processes with kinetic constants γ_1 and γ_2, respectively. The kinetic constant d_e is the external transport constant and γ_e is the degradation rate of external AHL. Note that in (3) we take into account the difference between the external and internal volumes with the factor $\frac{1}{N}$, as in [18].

In the model (1–3) we have made the following approximations: a) we assume the expression of mRNA is in a quasi-stationary state, neglecting the fast transient required by the mRNA concentration to reach its steady-state value, b) we consider the DNA/repressors complex also reaches very quickly its steady state value, allowing us to model the repression with a Hill function, c) we do not explicitly model the dimerization of the LuxR protein and its binding to AHL, and d) we assume that the LuxR gene is constitutively expressed and is not a limiting factor in the process.

For the purpose of obtaining analytic results, we model the variability between cells by taking the tightness of the PluxR promoter, κ^0, as a random variable with a normal distribution ($\kappa^0 \sim \mathcal{N}\left(\mu, \sigma^2\right)$). Here μ and σ^2 are the mean and variance across the population. However, to obtain more biologically-realistic results, more sources of variability should be included in the analysis. In the simulations we validated our analytical results by adding variability in all remaining parameters (we draw the parameters for each cell from a random distribution, see Fig. 4 and Section VI).

The following notation will be used hereafter: the partial states x_1 and x_2 are defined as $x_1 = \left[x_1^1, \ldots, x_1^N\right]^T \in \mathbb{R}^N$, $x_2 = \left[x_2^1, \ldots, x_2^N\right]^T \in \mathbb{R}^N$ and the full state $x = \left[x_1, x_2, x_e\right]^T \in \mathbb{R}^{2N+1}$. The vector of equilibrium points of the partial states x_1 and x_2 for the whole ensemble will be denoted $\bar{x}_1$ and $\bar{x}_2$. The equilibrium of x_e can be expressed as a function of the equilibrium

points of x_2^i:

$$\bar{x}_e = \frac{1}{N}\left(\frac{d_e}{d_e+\gamma_e}\right)\sum_i^N \bar{x}_2^i = \frac{\varepsilon}{N}\sum_i^N \bar{x}_2^i = \frac{\varepsilon}{N} 1_N^T \bar{x}_2, \qquad (4)$$

where $\varepsilon = \frac{d_e}{d_e+\gamma_e}$, and $\bar{x}_e, \bar{x}_2^i$ are the equilibrium values of x_e and x_2^i. The vector $1_N \in \mathbb{R}^N$ denotes the vector with all its elements equal to 1.

The variability of the promoter tightness κ^0 translates into a probability distribution for the steady state concentrations of LuxI and internal AHL across the population. In the remaining of the paper we will focus on quantifying this distribution, exploring how the intracellular feedback controller together with the cell-to-cell communication can be used to reduce variability of gene expression in the population.

III. Approximation of the Intracellular Controller

To simplify the analysis, we approximate the Hill function in (1) by the piecewise affine saturation function shown in Fig. 3. Its slope is taken to be the same as that of the Hill function at the half concentration constant θ, which is a sensible approximation for the typical values of the Hill coefficient n. Under this approximation, equation (1) can be rewritten as:

$$\dot{x}_1^i = u_{sat}(x_2^i) - \gamma_1 x_1^i, \qquad (5)$$

with

$$u_{\text{sat}}(x_2^i) = \begin{cases} \kappa_0^i + \kappa_1 & \text{if} \quad x_2^i < \theta - \delta \\ \kappa_0^i - \frac{\kappa_1 n}{4\theta} x_2^i + \frac{\kappa_1 n}{4} + \frac{\kappa_1}{2} & \text{if} \quad \left|x_2^i - \theta\right| < \delta \\ \kappa_0^i & \text{if} \quad x_2^i > \theta + \delta, \end{cases} \qquad (6)$$

and δ being the midpoint of the linear section:

$$\delta = \frac{2\theta}{n}. \qquad (7)$$

As mentioned before, the Hill function, together with its approximation u_{sat}, can be understood as a controller with fixed structure and tunable parameters. Our goal is to tune the controller parameters so as to shape the statistical distribution of the steady state concentration of AHL. We can conveniently reparameterize the saturation function u_{sat} using only three parameters. Rewriting (6) using $S = \frac{\kappa_1 n}{4\theta}$, $T = \theta$, $R = \frac{\kappa_1}{2}$ and $\delta = \frac{R}{S}$ we obtain

$$u_{sat}(x_2^i) = \begin{cases} \kappa_0 + 2R & \text{if} \quad x_2^i < T - \delta \\ \kappa_0 - S\left(x_2^i - T\right) + R & \text{if} \quad \left|x_2^i - T\right| < \delta \\ \kappa_0 & \text{if} \quad x_2^i > T + \delta. \end{cases} \qquad (8)$$

IV. Operation in the Linear Regime

In this section we obtain sufficient conditions under which the set

$$\Phi_x = \left\{x \in \mathbb{R}^{2N+1} : \left|\bar{x}_2^i - T\right| \le \delta, \forall i = 1, \ldots, N\right\}, \qquad (9)$$

is an attractive invariant set for all the cells in the interconnected population (see Fig. 3). Later in Section V we will use this result to obtain closed-form expressions for the mean and variance of the protein concentration.

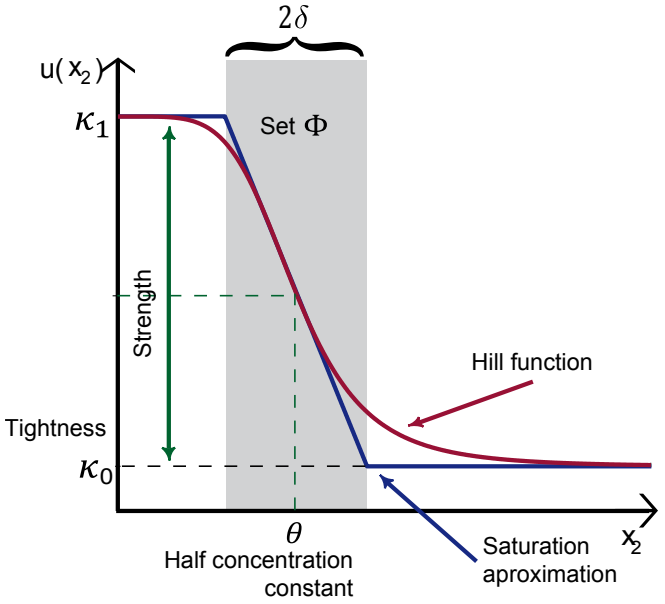

Fig. 3. Characteristic of the promoter. Hill function and its piecewise affine approximation for $n = 5$, $\theta = 5$ resulting in $\delta = 2$.

We first define the mean system states as $[\tilde{x_1}, \tilde{x_2}, x_e]^T$, with

$$\tilde{x}_j = \frac{1}{N}\sum_{i=1}^N x_i^j, \; j = 1, 2. \qquad (10)$$

The dynamics of the mean system can be expressed by

$$\dot{\tilde{x}}_1 = \rho - \gamma_1 \tilde{x}_1, \qquad (11)$$

$$\dot{\tilde{x}}_2 = \kappa_2 \tilde{x}_1 - (d+\gamma_2)\tilde{x}_2 + d x_e, \qquad (12)$$

$$\dot{x}_e = d_e \tilde{x}_2 - (d_e + \gamma_e) x_e. \qquad (13)$$

Where $\rho = \frac{1}{N}\sum_{i=1}^N u_{sat}(x_2^i)$. Notice that $\frac{1}{N}\sum_{i=1}^N \kappa_0^i = \tilde{\kappa}_0$ is the sample estimator of the mean μ, and that using (8) we get $\frac{1}{N}\sum_{i=1}^N u_{sat}(x_2^i) \in [\mu, 2R+\mu]$.

Equation (11) corresponds to an exponentially stable linear system with bounded input ρ and time constant $1/\gamma_1$. Therefore, $\tilde{x}_1$ there will get arbitrarily close to its steady state $\tilde{x}_1 \in \left[\frac{\mu}{\gamma_1}, \frac{\mu+2R}{\gamma_1}\right]$ after a finite time $t^\star$. The same argument holds for the subsystem formed by (12)-(13) with bounded input $\tilde{x}_1$, and thus x_e will get arbitrarily close to its steady state

$$x_e \in \frac{\varepsilon\kappa_2}{d(1-\varepsilon)+\gamma_2}\left[\frac{\mu}{\gamma_1}, \frac{\mu+2R}{\gamma_1}\right], \qquad (14)$$

after a finite time $t^{\star\star} > t^\star$.

Since x_e is a bounded signal, equations (2), (5) and (8) for the i^{th} cell can be rewritten using the variable change $z = x^i$:

$$\begin{cases} \dot{z}_1 = -\gamma_1 z_1 + u(z_2) \\ \dot{z}_2 = \kappa_2 z_1 - (d+\gamma_2) z_2 + z_e, \end{cases} \qquad (15)$$

where $z = [z_1, z_2]^T \in \mathbb{R}^2$ and $z_e = x_e$ is a bounded external perturbation with bounds $\underline{z}_e < z_e < \bar{z}_e$ from (14). Note that the index i was dropped for simplicity of notation.

The idea is to obtain a bound on the control signal $u(z_2)$ that ensures convergence of the trajectories of (15) to the set Φ, and that makes it invariant, with Φ defined as:

$$\Phi = \left\{ z \in \mathbb{R}^2 : \underline{z}_2 < z_2 < \overline{z}_2 \right\}. \tag{16}$$

To this end, it is convenient to use a structure for $u(z_2)$ which is more general than equations (6) and (8):

$$u(z_2) = \begin{cases} \kappa_1 + \kappa_0 & z_2 < \underline{z}_{2u} \\ u_{in}(z_2) & \underline{z}_{2u} < z_2 < \overline{z}_{2u} \\ \kappa_0 & z_2 > \overline{z}_{2u}. \end{cases} \tag{17}$$

with $\kappa_0 < u_{in}(z_2) < \kappa_1 + \kappa_0$ a bounded but not necessary continuous function of z_2 and $\underline{z}_2 < \underline{z}_{2u}$ and $\overline{z}_{2u} < \overline{z}_2$.

Note that making $T - \delta = \underline{z}_2$ and $T + \delta = \overline{z}_2$, equations (6) and (8) plus κ_0 represent (17), and when the set Φ is invariant, then the set Φ_x (9) is also invariant. Also, the set Φ can be rewritten as $\Phi = \underline{\Phi} \cap \overline{\Phi}$, the intersection of two sets $\underline{\Phi}$ and $\overline{\Phi}$ where

$$\underline{\Phi} = \left\{ z \in \mathbb{R} : \underline{\phi}(z_2) = -z_2 + \underline{z}_2 < 0 \right\} \tag{18}$$

and

$$\overline{\Phi} = \left\{ z \in \mathbb{R} : \overline{\phi}(z_2) = z_2 - \overline{z}_2 < 0 \right\}. \tag{19}$$

In [12] a boundary layer set is proven to be uniformly ultimately bounded, for systems with unitary relative degree (where the control action u appears explicitly in the first derivative of the output ϕ). The system in (15), however, has relative degree two when we take the control to be $u(z_2)$ and the outputs to be $\underline{\phi}(z_2)$ and $\overline{\phi}(z_2)$. To overcome this problem, we exploit the triangular structure of the system and design $u(z_2)$ using a backstepping-like approach. We can then use geometric invariance ideas to make the desired set Φ invariant. With this technique we get the following result.

Theorem 1: The set Φ is an invariant and attractive set for z_2 if the following inequalities hold

$$\kappa_0 + \kappa_1 > \gamma_1 \frac{(d+\gamma_2)\underline{z}_2 - \underline{z}_e}{\kappa_2} \tag{20}$$

and

$$\kappa_0 < \gamma_1 \frac{(d+\gamma_2)\overline{z}_2 - \overline{z}_e}{\kappa_2} \tag{21}$$

Proof: The proof will be sketched in four steps, two for the set $\underline{\Phi}$ (18) and another two for the set $\overline{\Phi}$ (19). First we will use $z_1 = u^\star$ as a virtual control action in the second equation of system (15). This will allow us to find the set $\underline{\mathscr{S}}$ of all z_1 that make $\underline{\Phi}$ an invariant set for z_2. In the second design step, we will find the set of all $u(z_2)$ that make $\underline{\mathscr{S}}$ invariant for z_1.For the remaining two steps we proceed in a similar way to make $\overline{\mathscr{S}}$ and $\overline{\Phi}$ invariant sets for z_1 and z_2 respectively.

First consider

$$\begin{cases} \dot{z}_2 = \kappa_2 u^\star - (d+\gamma_2) z_2 + z_e \\ \underline{\phi}(z_2) = -z_2 + \underline{z}_2 \end{cases} \tag{22}$$

And now the goal is to find $u^\star$ in order to make set $\underline{\Phi}$ invariant.

$$\underline{\Phi} = \left\{ z_2 \in \mathbb{R}^+ : \underline{\phi}(z_2) < 0 \right\} \tag{23}$$

This is equivalent to find bounds on $u^\star$ to make the vector field point inside $\underline{\Phi}$, when z_2 reaches the boundary of the set $\underline{\Phi}$.

Consider the z_2-system (22). Using $f(z_2, z_e) = -(d+\gamma_2)z_2 + z_e$ and $g(z_2) = \kappa_2$ it is possible to rewrite $\dot{z}_2 = f(z_2, z_e) + g(z_2)u^\star$. In order to find the lower bound $u^\star$, we need to find the direction of the vector field with respect to the boundary of the set $\underline{\Phi}$. To obtain the control signal that makes $\underline{\Phi}$ invariant, we apply the explicit invariance condition [13]:

$$u \begin{cases} \leq u^\phi & : z \in \partial\Phi \wedge L_g\phi > 0 \\ \geq u^\phi & : z \in \partial\Phi \wedge L_g\phi < 0 \\ \text{not defined} & : z \in \partial\Phi \wedge L_g\phi = 0 \\ \text{free} & : z \in \Phi \setminus \partial\Phi \end{cases} \tag{24}$$

Here $L_f\phi$ accounts for the Lie derivative of ϕ along the direction of the vector field f, and $u^\phi = -\frac{L_f\phi}{L_g\phi}$.

In our case, we get

$$u^{\star\underline{\phi}} = -\frac{(d+\gamma_2)z_2 - z_e}{-\kappa_2} \tag{25}$$

and hence $\underline{z}_2$ is an upper bound of z_2. For the perturbation term z_e, we will use its lower bound $\underline{z}_e$, as it is an anti-cooperative term with respect to z_2. We obtain the following lower bound for $u^\star$, that we will call $u^\star_{min}$:

$$u^\star_{min} = \frac{d+\gamma_2\underline{z}_2 - \underline{z}_e}{\kappa_2} \tag{26}$$

Then, if the following holds

$$u^\star > u^\star_{min}, \quad \forall z_2 \notin \underline{\Phi}. \tag{27}$$

the trajectory of z_2 is forced to remain inside $\underline{\Phi}$.

Secondly, we proceed with the first equation of system (15). Remembering that $u^\star = z_1$, we have

$$\begin{cases} \dot{z}_1 = -\gamma_1 z_1 + u(z_2) \\ \underline{\sigma}(z_1) = -z_1 + u^\star_{min} \end{cases} \tag{28}$$

Define the set $\underline{\mathscr{S}}$ as:

$$\underline{\mathscr{S}} = \left\{ z_1 \in \mathbb{R}^+ : \underline{\sigma}(z_1) < 0 \right\} \tag{29}$$

when $z_2 \notin \underline{\Phi}$. The goal now is to find a bound on $u(z_2)$ so as to make the set $\underline{\mathscr{S}}$ invariant.

Using the same methodology and the explicit invariant condition, we get

$$u^{\underline{\sigma}} = \gamma_1 z_1. \tag{30}$$

To bound z_1 we use the fact that the application of the control signal (30) is only required when $\underline{\sigma} > 0$. Hence $u^\star_{min}$ is an upper bound of z_1. This allows us to obtain the bound u_{min} for u:

$$u_{min} = \gamma_1 u^\star_{min} = \gamma_1 \frac{(d+\gamma_2)\underline{z}_2 - \underline{z}_e}{\kappa_2} \tag{31}$$

Then the set $\underline{\Phi}$ is an invariant and attractive set for z_2 if

$$u(z_2) > \gamma_1 \frac{(d+\gamma_2)\underline{z}_2 - \underline{z}_e}{\kappa_2} \quad \forall z_2 \notin \underline{\Phi} \tag{32}$$

Equivalent steps can be applied with the set $\overline{\Phi}$ to obtain:

$$u(z_2) < \gamma_1 \frac{(d+\gamma_2)\bar{z}_2 - \bar{z}_e}{\kappa_2} \quad \forall z_2 \notin \overline{\Phi}. \tag{33}$$

Now, using $u(z_2) = \kappa_0$ when $z_2 \notin \overline{\Phi}$, and $u(z_2) = \kappa_0 + \kappa_1$ when $z_2 \notin \underline{\Phi}$, we get that the following inequalities

$$\kappa_0 + \kappa_1 > \gamma_1 \frac{(d+\gamma_2)\underline{z}_2 - \underline{z}_e}{\kappa_2} \tag{34}$$

and

$$\kappa_0 < \gamma_1 \frac{(d+\gamma_2)\bar{z}_2 - \bar{z}_e}{\kappa_2} \tag{35}$$

must hold to make Φ an invariant set. ■

V. Control of protein mean and variance

In this section we show that the mean and variance of the AHL concentration across the cell population can be controlled independently with different parameters of the controller. This constitutes the main contribution of this paper.

Under the conditions of Theorem 1 we know that every individual system will eventually operate in the linear regime:

$$\dot{x}_1^i = \kappa_0^i - S\left(x_2^i - T\right) + R - \gamma_1 x_1^i, \tag{36}$$

The dynamics of the whole ensemble can then be written as a $(2N+1)-$dimensional linear system:

$$\dot{x} = \left[\begin{array}{cc|c} -\gamma_1 \mathbf{I}_N & S\mathbf{I}_N & 0_N \\ \kappa_2 \mathbf{I}_N & -(d+\gamma_2)\mathbf{I}_N & d1_N \\ \hline 0_N^T & \frac{d_e}{N} 1_N^T & -(d_e+\gamma_e) \end{array}\right] x + \left[\begin{array}{c} \kappa_0 + (ST+R)\,1_N \\ 0_N \\ \hline 0 \end{array}\right], \tag{37}$$

We define the matrix Π_N as

$$\Pi_N = \mathbf{I}_N - \frac{1}{N}\mathbf{1}_{N\times N}, \tag{38}$$

where $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the identity matrix and $\mathbf{1}_{N\times N} \in \mathbb{R}^{N\times N}$ has all its entries equal to 1.

Note that Π_N is idempotent (*i.e.* $\Pi_N\Pi_N = \Pi_N$) and satisfies $\Pi_N 1_N = 0_N$.

Setting $\dot{x} = 0$ in (37) and using (4) we get a system of $2N$ linear equations for the steady states $\bar{x}_1$ and $\bar{x}_2$:

$$\left[\begin{array}{cc} \gamma_1 \mathbf{I}_N & S\mathbf{I}_N \\ \kappa_2 \mathbf{I}_N & -[d(1-\varepsilon)+\gamma_2]\mathbf{I}_N - d\varepsilon\Pi_N \end{array}\right] \left[\begin{array}{c} \bar{x}_1 \\ \bar{x}_2 \end{array}\right] = \left[\begin{array}{c} \kappa_0 + (ST+R)\,1_N \\ 0_N \end{array}\right] \tag{39}$$

Theorem 2: Under the conditions for Theorem 1, the mean and variance of the distribution of x_2 for a population of N cells can be controlled independently by tuning the parameters of each cell intracellular controller as follows:

$$\mathrm{E}\left\{\bar{x}_2^i\right\} = \frac{1}{S+\beta}\left(\mu + ST + R\right) \tag{40}$$

$$\mathrm{Var}\{\bar{x}_2^i\} = \left[\frac{1}{(S+\beta-\varepsilon d)^2} + \frac{1}{N}\left(\frac{1}{(S+\beta)^2} - \frac{1}{(S+\beta-\varepsilon d)^2}\right)\right]\sigma^2 \tag{41}$$

Proof: Consider the system in (2), (3) and (5) with the approximation in (8) under the conditions of Theorem 1, then all cells in the population operate in the linear region Φ. From (39) we have

$$\gamma_1 \bar{x}_1 + S\bar{x}_2 = \kappa_0 + (ST+R)\,1_N \tag{42}$$

and

$$\bar{x}_1 = \left[(d(1-\varepsilon)-\gamma_2)\,\mathbf{I_N} + d\varepsilon\Pi_N\right]\frac{\bar{x}_2}{\kappa_2}. \tag{43}$$

Replacing (43) in (42), we obtain

$$\left[\left(S + \frac{[d(1-\varepsilon)+\gamma_2]\,\gamma_1}{\kappa_2}\right)\mathbf{I_N} + d\varepsilon\Pi_N\right]\bar{x}_2 = \kappa_0 + (ST+R)\,1_N. \tag{44}$$

Exploiting the structure of the matrix in (44) we obtain a closed-form expression for $\bar{x}_2$ (details omitted for brevity):

$$\bar{x}_2 = \left[\frac{1}{\alpha}\mathbf{I_N} + \left(\frac{1}{\alpha-\varepsilon d} - \frac{1}{\alpha}\right)\Pi_N\right]\left(\kappa_0 + (ST+R)\,1_N\right), \tag{45}$$

where $\beta = \frac{[d(1-\varepsilon)+\gamma_2]\gamma_1}{\kappa_2}$ and $\alpha = S+\beta$.

If $\kappa^0 \sim \mathcal{N}\left(\mu, \sigma^2\right)$, the expected value of $\bar{x}_2$ along the population can be computed using properties of the Π_N matrix:

$$\mathrm{E}\{\bar{x}_2\} = \frac{1}{S+\beta}\left(\mu + ST + R\right)1_N \tag{46}$$

A closed-form expression for the variance of x_2 can be obtained similarly

$$\mathrm{Var}\{\bar{x}_2\} = \left[\frac{1}{\alpha^2}\mathbf{I_N} + \left(\frac{1}{(\alpha-\varepsilon d)^2} - \frac{1}{\alpha^2}\right)\Pi_N\right]\sigma^2\mathbf{I_N}. \tag{47}$$

The expressions in (40)–(41) can be finally obtained directly from (46)–(47).

■

The expression in (41) indicates that the variance can be controlled independently from the mean with the parameter S, and its sensitivity is

$$\frac{\partial \mathrm{Var}\{\bar{x}_2^i\}}{\partial S} = -\frac{2}{N}\left[\frac{1}{(S+\beta)^3} + \frac{N-1}{N(S+\beta-\varepsilon d)^3}\right]\sigma^2, \tag{48}$$

which indicates that a steep feedback (i.e. with a high value of S) tends to reduce the population level variability.

VI. Simulations

To demonstrate the potential of the proposed control strategy, we ran numerical simulations of the different control architectures shown in Fig. 1. The parameters values used are shown in Table I. These are physiologically realistic values for *E. coli*, and similar to those typically found in the literature [15], [16], [18]–[21].

In order to compare the different architectures, we observe the variances and adjust the means to be similar in all cases.

Note that nowadays, the Ribosome Binding Site (RBS) strength is one of the more suitable biological tuning knobs, which can be selected with high predictability [22]. Thus,

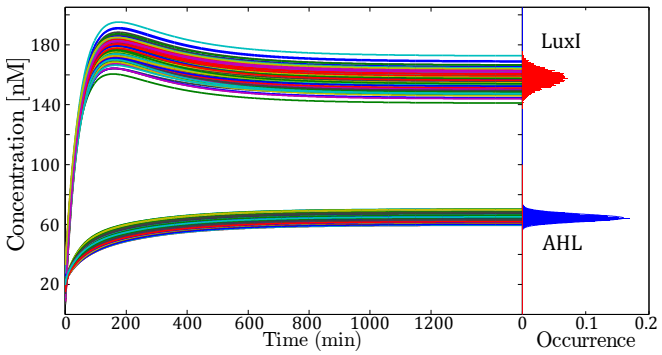

Fig. 4. Time course and steady state histogram of 10.000 cells for architecture D with closed loop with cell to cell communication (with $n = 1.5$)

consider rewriting (1) taking into account the RBS strength in the following way:

$$\dot{x}_1^i = \mathrm{RBS}_s \left(\hat{\kappa}_0^i + \hat{\kappa}_1^i \frac{\theta^n}{\theta^n + x_2^{i\,n}} \right) - \gamma_1 x_1^i. \qquad (49)$$

The new parameters $\hat{\kappa}_0$ and $\hat{\kappa}_1$ are fixed among all the examples to a 10% of leakiness and 10-fold dynamic range [18], [19], [21]. Hence, the RBS strength RBS_s will be used as the only tuning knob in order to have the same mean of AHL in architectures B, C and D than in architecture A.

Fig. 4 shows the simulation of 10000 ODEs (equations 1-2) one for each cell i. The values for the parameters were drawn from a normal distribution with mean equal to the nominal value of the parameter in Table I and a variance of 5% of that value. The steady state distributions for all the architectures (see Fig. 5) were obtained from the same kind of simulations with corresponding parameters (Table I).

In Fig. 5 we can see how the different architectures impact on the steady state distribution of AHL. The mean, the variance, and the coefficient of variation (CV) of AHL resulting from the simulations are shown also in Table I.

From the results in Table I and from the distributions of AHL in Fig. 5, it appears that both the feedback-regulated production of LuxI, and the cell-to-cell communication through AHL are required for best performance. Case D (bottom plot from Fig. 5) have smaller variances than cases A, B and C (top plot from Fig. 5).

Also comparing the two distributions in the bottom plot from Fig. 5 it appears that a steep feedback $n = 3$ tends to reduce the population level variability as predicted by (48) with respect to a less steeper one $n = 1.5$.

VII. Conclusions

In this paper, we investigated the design of a synthetic gene controller aimed at reducing gene expression variability at the population level. As a proposed synthetic biology implementation, we considered a cell-to-cell communication system coupled with an intracellular genetic controller characterized by a sigmoidal nonlinearity. To simplify the mathematical analysis, the nonlinearity was approximated by a piecewise linear function. Based on this approximation,

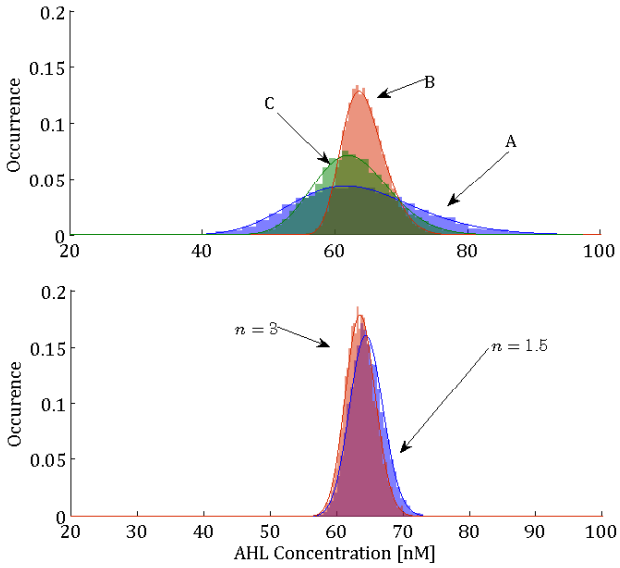

Fig. 5. Steady state distributions for 10.000 cells of AHL concentration in the different architectures of Fig. 1. In the top figure architectures A, B and C are shown. The bottom figure shows, architecture D with $n = 1.5$ and also with a steeper controller $n = 3$.

TABLE I

Nominal values for the parameters used in the simulations. Mean, variance and CV for each architecture.

	Architecture						
Parameter	A	B	C	D		Units	Reference
γ_1	0.0173					min^{-1}	[20,21]
RBS_s	0.76	15	0.125	4.6	4.6	nM.min^{-1}	[19,21]
θ	-	-	63.24			nM	[15,21]
n	-	-	1.5	1.5	3	-	[15]
κ_2	0.04					min^{-1}	[21]
d	-	0.3	-	0.3		min^{-1}	[16]
d_e	-	0.006	-	0.006		min^{-1}	[16]
γ_2	2.82e-3					min^{-1}	[18,20]
$\mathrm{E}\{x_2^i\}$	63.55	64.36	62.65	63.25	63.25	nM	
$\mathrm{Var}\{x_2^i\}$	89.43	10.10	32.13	6.23	4.12	nM	
$\mathrm{CV}\{x_2^i\}$	0.1488	0.0494	0.0905	0.0395	0.0321		

we established: (i) conditions under which the non saturated region of the controller is an attractive invariant set, and (ii) closed-form expressions for the first two moments of the distribution of AHL across a population. We also demonstrated how the parameters of the controller can be fine-tuned to independently control the mean and the variance of the distribution.

With the progress of current experimental techniques, adjusting genetic parameters has become feasible and has paved the way for the use of rigorous control-theoretic approaches for the design of genetic circuits. In this line we think that having a model-based guideline to design genetic networks has tremendous potential in Synthetic Biology.

For example, changing cooperativity with protein sequestration techniques [23] can allow us to tune S (the slope of the nonlinearity), whereas sequence repeats in the spacer region of the RBS [22] can be used to adjust R by changing the RBS strength. The threshold T can also be tuned with similar techniques [23], for example by shifting the position where the complex LuxR-AHL binds (called Lux-box) [15] or by making single point mutations in the Lux-box.

As a proof of concept, in this work we proposed the basis for protein distribution control along a population by controlling the distribution of the signalling molecules AHL. Using this signal to drive the production of a protein of interest could be used to control its distribution. We are working in that direction and investigating different ways in which this could be done. We are also considering the implementation of the proposed genetic circuit *in vivo* using synthetic biological parts (for example Biobricks), and the analysis of mixed population scenarios in which feedback-regulated cells coexist with unregulated ones, so as to design distributed approaches to biocomputing.

REFERENCES

[1] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, Jan. 2000.

[2] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, Jan. 2000.

[3] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems." *Nature reviews. Molecular cell biology*, vol. 10, no. 6, pp. 410–22, Jun. 2009.

[4] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age." *Nature reviews. Genetics*, vol. 11, no. 5, pp. 367–79, May 2010.

[5] D. McMillen, N. Kopell, J. Hasty, and J. Collins, "Synchronizing genetic relaxation oscillators by intercell signaling," *Proceedings of the National Academy of Sciences*, vol. 99, no. 2, p. 679, 2002.

[6] H. Kobayashi, M. Kaern, M. Araki, K. Chung, T. Gardner, C. Cantor, and J. Collins, "Programmable cells: interfacing natural and engineered gene networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 22, p. 8414, 2004.

[7] S. Basu, Y. Gerchman, C. Collins, F. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, no. 7037, pp. 1130–1134, 2005.

[8] F. K. Balagaddé, H. Song, J. Ozaki, C. H. Collins, M. Barnet, F. H. Arnold, S. R. Quake, and L. You, "A synthetic Escherichia coli predator-prey ecosystem." *Molecular systems biology*, vol. 4, p. 187, Jan. 2008.

[9] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic Gene Expression in a Single Cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, Aug. 2002.

[10] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, "Moment-Based Inference Predicts Bimodality in Transient Gene Expression," *PNAS*, vol. 109, no. 21, pp. 8340–8345, 2012.

[11] P. S. Swain, M. B. Elowitz, and E. D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12 795–800, Oct. 2002.

[12] F. Esfandiari and H. Khalil, "Stability analysis of a continuous implementation of variable structure control," *Automatic Control, IEEE Transactions on*, vol. 36, no. 5, pp. 616–620, 1991.

[13] A. Vignoni, J. Picó, F. Garelli, and H. De Battista, "Sliding mode reference conditioning for coordination in swarms of non-identical multi-agent systems." in *Variable Structure Systems, 12th IEEE International Workshop on*, 2012, pp. 231–236.

[14] C. Fuqua, M. Parsek, and E. Greenberg, "Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing," *Annual review of genetics*, vol. 35, no. 1, pp. 439–468, 2001.

[15] K. A. Egland and E. P. Greenberg, "Conversion of the vibrio fischeri transcriptional activator, LuxR, to a repressor," *Journal of Bacteriology*, vol. 182, no. 3, pp. 805–811, Feb. 2000.

[16] H. Kaplan and E. Greenberg, "Diffusion of autoinducer is involved in regulation of the vibrio fischeri luminescence system." *Journal of Bacteriology*, vol. 163, no. 3, pp. 1210–1214, 1985.

[17] A. L. Schaefer, D. L. Val, B. L. Hanzelka, J. E. Cronan, and E. P. Greenberg, "Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified vibrio fischeri luxi protein," *Proceedings of the National Academy of Sciences*, vol. 93, no. 18, pp. 9505–9509, 1996.

[18] P. Mina, M. di Bernardo, N. J. Savery, and K. Tsaneva-Atanasova, "Modelling emergence of oscillations in communicating bacteria: a structured approach from one to many cells," *Journal of The Royal Society Interface*, vol. 10, no. 78, 2013.

[19] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: from theories to phenotypes," *Nature Reviews Genetics*, vol. 6, no. 6, pp. 451–464, 2005.

[20] C. Smith, H. Song, and L. You, "Signal discrimination by differential regulation of protein stability in quorum sensing," *Journal of molecular biology*, vol. 382, no. 5, p. 1290, 2008.

[21] M. Weber, J. Buceta *et al.*, "Dynamics of the quorum sensing switch: stochastic and non-stationary effects," *BMC systems biology*, vol. 7, no. 1, p. 6, 2013.

[22] R. G. Egbert and E. Klavins, "Fine-tuning gene networks using simple sequence repeats," *Proceedings of the National Academy of Sciences of the United States of America*, Aug. 2012.

[23] N. E. Buchler and F. R. Cross, "Protein sequestration generates a flexible ultrasensitive response in a genetic network," *Molecular Systems Biology*, vol. 5, 2009.