Universal structural requirements for maximal robust perfect adaptation in biomolecular networks

Ankit Gupta and Mustafa Khammash

ETH Zürich, Department of Biosystems Science and Engineering,
Mattenstrasse 26, 4058 Basel, Switzerland

Abstract

Consider a biomolecular reaction network that exhibits robust perfect adaptation to disturbances from several parallel sources. The well-known *Internal Model Principle* of control theory suggests that such systems must include a subsystem (called the “internal model”) that is able to recreate the dynamic structure of the disturbances. This requirement poses certain structural constraints on the network which we elaborate in this paper for the scenario where constant-in-time disturbances maximally affect network interactions and there is model uncertainty and possible stochasticity in the dynamics. We prove that these structural constraints are primarily characterized by a simple linear-algebraic stoichiometric condition which remains the same for both deterministic and stochastic descriptions of the dynamics. Our results reveal the essential requirements for maximal robust perfect adaptation in biology, with important implications for both systems and synthetic biology. We exemplify our results through many known examples of robustly adapting networks and we construct new examples of such networks with the aid of our linear-algebraic characterization.

Keywords: Robust Perfect Adaptation; RPA; Reaction Networks; Systems Biology; Synthetic Biology.

1 Introduction

The phenomenon of adaptation is central to biology [1]. Nearly all living systems exhibit some form of adaptation — almost by necessity — as they need to survive and maintain their functions in harsh and unpredictable environments [2]. Also known as homeostasis in biology [3], the property of adaptation refers to the ability of a system to counteract persistent external disturbances and closely maintain some key component (called the “output”) of its internal state. If there is no difference in the steady-state level of this output (called the “set-point”), pre- and post-disturbance, then the system is said to be perfectly adapting. If this stringent form of adaptation holds in spite of perturbations to the parameters or the structure of the network underlying the system, then this property is called robust perfect adaptation (RPA). Biological examples of RPA systems include chemotaxis in prokaryotes [4] and eukaryotes [5], intracellular

osmolarity [6], stress response mechanisms [7], transcription factor σ -70 activity [8], ammonia uptake [9], copper, zinc, iron and calcium homeostasis [9, 10, 11], glucose uptake [12] and pattern formation mechanisms during development [13].

Evidently the notion of RPA is fairly nuanced and its precise mathematical definition presupposes many specifications, such as how the set-point for the output species is encoded in terms of network parameters, what type of disturbances can affect the network, and how these effects are manifested by the network dynamics. Furthermore this dynamics can have the classical deterministic description or a stochastic description [14] — the latter being more suitable when the network involves biomolecular species with low abundance levels [15]. Existing quantitative studies on RPA systems have mainly focussed on deterministic single-input single-output (SISO) networks where the disturbances can only enter the network dynamics via a pre-specified input node [16, 17, 18, 19]. Even though these SISO networks are quite popular in many engineering domains, the single-input assumption seems unrealistic in biology as disturbances that arise are typically very complex and do not have a single well-defined entry point. In the context of intracellular networks, such disturbances include cross-talk between networks [20], burden effects [21, 22], context dependence [23], unmodelled interactions and a myriad of other extrinsic factors [1]. These disturbances can simultaneously alter several networks parameters that influence the reaction rates, and they can also modify the network structure by adding or removing reaction components. As many of these disturbances cannot be avoided or even mitigated, it makes strong evolutionary sense for intracellular RPA networks to attain a form of *maximal robustness* in which the set-point for the key output variable depends on the least number of network parameters, and it is insensitive to *all* the others. Our aim in this paper is to mathematically characterise and study the structure of biomolecular networks that constitute such maximal RPA (or *maxRPA*) systems.

The scope of our work is very different from existing studies that seek to topologically characterise RPA networks [18, 19]. The primary difference is that we work with a stronger notion of robustness and consequently our characterising conditions for RPA networks are more restrictive than the topological requirements expounded in these studies. In particular our conditions pertain to the stoichiometric structure of a RPA network, and they are both necessary and sufficient for the RPA property to hold and the output set-point to remain invariant to the *largest possible* set of network disturbances. On the other hand, the conditions on the topological structure specified in [18, 19] are only necessary and they only guarantee insensitivity of the output set-point to the designated input parameter, but this set-point can arbitrarily depend on other network parameters. Moreover, as the topological structure is determined by the signs of the network interactions at the steady-state, it is conceivable that this structure depends non-trivially on the choice of network parameters, further complicating the issue of existence of the RPA property. The other main difference between our work and these existing RPA studies [18, 19] is that in addition to the deterministic setting, we consider the stochastic setting as well. This allows us to delineate the role of dynamical randomness in influencing the RPA requirements.

Motivated by applications in synthetic biology, there have been many exciting developments in recent years in the area of designing biomolecular controller modules that can be interfaced with arbitrary intracellular networks to achieve RPA for some output species of interest. Some of the controller modules that have been proposed and

Figure 1: **Schematic of the internal model principle (IMP):** A RPA network decomposes into an *internal model (IM)* and the *rest of the network (RON)*. Robustness to disturbances is achieved by a feedback mechanism that is realised by the IM. The IM dynamically measures the disturbances via the deviation of the regulated output X_1 from the set-point, and then passes *actuating* signals to the RON to counter the disturbances.

validated, either theoretically and/or experimentally, are the antithetic integral feedback (AIF) controller [8, 24], the autocatalytic controller [25, 26] and the zeroth-order degradation controller [10]. The design paradigm of synthetic biology naturally imposes a clear separation between the controller and the “controlled” networks, which cannot be *a priori* assumed when one studies endogenous RPA networks from the standpoint of systems biology. However a deep result from control theory, called the *internal model principle (IMP)*, stipulates that every RPA network can, at least theoretically, be separated into two components - an *internal model (IM)* that dynamically evaluates the effect of disturbance by measuring the regulated output, and generates counteracting signals, that are then passed to the *rest of the network (RON)* via the actuation mechanism to complete the feedback loop (see Figure 1). Importantly, the IMP states that in order to reject constant-in-time disturbances, the actuation signal computed by IM must measure the time-integral of the *error* signal, which is the difference between the regulated output and its set-point. Hence the IMP should facilitate implementation of the famed *integral feedback* mechanism in control theory that has numerous engineering applications. In the setting of synthetic biology, the controller module (like the AIF, the autocatalytic or the zeroth-order degradation controller) renders the IMP, while the controlled network becomes RON. It is important to point out that IMP does not refer to a single result but rather it provides a template for many possible results; each of which will hold under specific assumptions about the network model and the class of disturbances which are rejected. For example, the IMP has been proved for linear systems [27, 28] and for a certain nonlinear systems where disturbances enter the dynamics linearly [29]. In this paper we provide versions of the IMP in both deterministic and stochastic settings, which are tailored for biomolecular reaction networks exhibiting maxRPA. This allows us to decompose each maxRPA network into an IM and the RON, as in Figure 1, and we explicitly identify an integrator

that provides the integral action necessary for robustness.

Our main characterisation results posit that maxRPA networks are fully characterised by certain stoichiometric conditions that are simple to state and verify. The principal requirement for maxRPA is a linear-algebraic condition on the network’s stoichiometry, which *remains the same* in both deterministic and stochastic settings (see Sections 3.1 and 3.2). The only difference between the two scenarios is in the requirements for the stoichiometries of the reactions, with the mass-action form [14], whose rate constants are assumed to encode the output set-point. We explicitly identify the set-point encoding function and show that maxRPA characterising conditions are more restrictive in the stochastic case. It follows that stochastic maxRPA networks form a strict subset of deterministic maxRPA networks, which extends a similar result proved in the context of synthetic biology in [24]. Interestingly, our results show that randomness in the dynamics necessitates that the maxRPA network must embed the *antithetic* mechanism, which is characterised by the presence of a sequestration or an annihilation reaction between two IM species (see Section 3.3). We refer to maxRPA networks without such a mechanism as *homothetic* and they can only exist in the deterministic setting. For any maxRPA network we can view its IM as an inbuilt controller that generates robustness for the regulated output variable. The metabolic cost incurred by this controller can be adjudged by the complexity of the IM, in terms of the number of species it contains and the number of reactions among them. It is natural to ask — what are the least complex IMs that are able to confer robustness? We examine this question in both deterministic and stochastic settings and fully characterise the structure of all maxRPA networks with minimal IMs (see Section 4).

In Section 5 we consider known biological examples from the literature and show how they fit our results. These include the bacterial chemotaxis network [4, 30], the network governing proliferative control of cell lineages [31] and the quasi-integral controller based on phosphorylation-dephosphorylation cycles [32]. We also discuss the *sniffer system* [33], which utilises an incoherent feedforward loop (IFFL) to achieve RPA w.r.t. to some of the network parameters, but is not a maxRPA network. Finally in Section 6 we conclude with possible implications of our results.

2 Preliminaries

In this section we provide descriptions of reaction networks in both deterministic and stochastic settings. Then we mathematically define our criterion for maxRPA and discuss how this property is connected to the existence of *integrators* [2].

2.1 Reaction Networks

Consider a reaction network with N biomolecular species $\mathbf{X}_1, \dots, \mathbf{X}_N$ that interact among each other through K reactions of the form

$$\sum_{i=1}^N \nu_{ik} \mathbf{X}_i \longrightarrow \sum_{i=1}^N \nu'_{ik} \mathbf{X}_i \quad (2.1)$$

where ν_{ik} (resp. ν'_{ik}) are *stoichiometric* constants in $\mathbb{N}_0$ (i.e. the set of nonnegative integers) denoting the number of molecules of species $\mathbf{X}_i$ that are consumed

(resp. produced) by reaction k . Hence the change in the vector of species molecular counts, brought about by a firing of reaction k , is given by the *stoichiometric vector* $\zeta_k = (\nu'_{1k} - \nu_{1k}, \dots, \nu'_{Nk} - \nu_{Nk}) \in \mathbb{Z}^N$ (where $\mathbb{Z}$ is the set of all integers).

Deterministic model: In the classical deterministic model of a reaction network, the state $x = (x_1, \dots, x_N) \in \mathbb{R}_+^N$ (where $\mathbb{R}_+$ is the set of nonnegative reals) denotes the vector of species concentrations. For each reaction k , the state x is mapped to the *flux* for reaction k according to a *propensity function* $\lambda_k : \mathbb{R}_+^N \rightarrow \mathbb{R}_+$. Then the reaction dynamics $(x(t))_{t \geq 0}$ is described by the following system of ODEs

$$\frac{dx}{dt} = D(x) := \sum_{k=1}^K \lambda_k(x) \zeta_k. \quad (2.2)$$

For our results, we shall be assuming that a solution of this system of ODEs exists uniquely for some open set $\mathcal{E} \subset \mathbb{R}_+^N$ which is time-invariant under the dynamics, and there is a globally attracting fixed point $\bar{x}$, satisfying

$$D(\bar{x}) = \mathbf{0}. \quad (2.3)$$

If the propensity function for reaction k has *mass-action* form, then it means that λ_k is given by

$$\lambda_k(x) = \theta_k \prod_{i=1}^N x_i^{\nu_{ik}} \quad (2.4)$$

where θ_k is a positive reaction rate constant and ν_{ik} -s are as in (2.1).

Stochastic model: In the standard stochastic model of a reaction network, the state $x = (x_1, \dots, x_N) \in \mathbb{N}_0^N$ denotes the vector of species copy-numbers or molecular counts. The propensity function $\lambda_k : \mathbb{N}_0^N \rightarrow \mathbb{R}_+$ maps the state x to the rate of firing $\lambda_k(x)$ of reaction k , and when this reaction fires it displaces the state to $(x + \zeta_k)$ where ζ_k is the stoichiometry vector as before. It is natural to impose the following restriction on each propensity function λ_k

$$\lambda_k(x) > 0 \quad \iff \quad x \geq \nu_k \quad (\text{componentwise}), \quad (2.5)$$

that essentially means that each reaction k has a positive rate of firing if and only if the requisite number of reactant molecules (given by ν_{ik} for species $\mathbf{X}_i$) are present. This assumption is biologically reasonable and it ensures that the non-negative integer orthant $\mathbb{N}_0^N$ is invariant for the stochastic dynamics. This assumption is trivially satisfied for the propensity function λ_k if it has the *mass-action* [14] form

$$\lambda_k(x_1, \dots, x_N) = \theta_k \prod_{i=1}^N \frac{x_i(x_i - 1) \dots (x_i - \nu_{ki} + 1)}{\nu_{ki}!}, \quad (2.6)$$

where θ_k is a positive reaction rate constant as before.

In the stochastic model, the reaction dynamics $(X(t))_{t \geq 0}$ is described by a *continuous-time Markov chain* (CTMC) [14]. Letting $\mathcal{E} \subset \mathbb{N}_0^N$ to be the set of all accessible states for the CTMC and p_t to be probability distribution of $X(t)$ i.e.

$$p_t(x) = \mathbb{P}(X(t) = x), \quad x \in \mathcal{E},$$

the well-known *Chemical Master Equation* (CME) [14] expresses the time-evolution of p_t as

$$\frac{dp_t(x)}{dt} = \sum_{k=1}^K (\lambda_k(x - \zeta_k)p_t(x - \zeta_k) - \lambda_k(x)p_t(x)) \quad \text{for all } x \in \mathcal{E}. \quad (2.7)$$

The stability of the CTMC $(X(t))_{t \geq 0}$ is given by the notion of ergodicity which mandates that regardless of the initial distribution $p_0(\cdot)$, as $t \rightarrow \infty$, the probability distribution $p_t(\cdot)$ converges to a stationary distribution π which is a globally attracting fixed point for the CME (2.7). If this convergence is exponentially fast then this property is called exponential ergodicity. In order to verify ergodicity, a two-step approach is needed. First one needs to check if the state-space $\mathcal{E}$, which is typically infinite, is *irreducible*¹ and then construct a suitable Foster-Lyapunov function² over the state-space $\mathcal{E}$ [34]. For stochastic reaction networks, computational frameworks for systematically carrying out both these tasks are provided in [35] and [36] respectively.

2.2 The maxRPA property

Suppose that the reaction structure, given by reactions (2.1), is fixed. Henceforth we designate the first species $\mathbf{X}_1$ as the *output* species. The property of RPA pertains to robustly driving the steady-state abundance level of the output species to some set-point θ^* , despite *perturbations* or *disturbances* to the network. The steady-state of the network, in both deterministic and stochastic settings, is determined by the propensity functions, which may depend on parameters such as reaction rate constants for mass-action reactions or other parameters, like temperature, which influence the kinetics. It is natural to assume that the set-point is encoded as a function of these parameters, and for definiteness, we suppose that the first m reactions in the network have mass-action kinetics and their rate constants $\theta_1, \dots, \theta_m$ encode the set-point θ^* via some smooth function ϕ , i.e.

$$\phi(\theta_1, \dots, \theta_m) = \theta^*.$$

Noting that this set-point will remain invariant to time-rescaling, one can derive a first-order *partial differential equation* (PDE) for ϕ which shows that m must at least be two (see Section S2 in the Supplement). As we are concerned with maxRPA, we would like our set-point to be determined by the least number of parameters. Hence we set $m = 2$ and in this case our PDE shows that ϕ is essentially only a function of the ratio of the parameters θ_1 and θ_2 . Denoting this function as ϕ_{out} , the set-point becomes

$$\theta^* = \phi_{\text{out}}\left(\frac{\theta_1}{\theta_2}\right) \quad \text{with} \quad \phi'_{\text{out}}\left(\frac{\theta_1}{\theta_2}\right) \neq 0. \quad (2.8)$$

The non-zero derivative condition is added to rule out the trivial case where the set-point encoding function does not depend on its argument. Our results will show that

¹The irreducibility condition demands that for any two states $x, y \in \mathcal{E}$ there must exist a positive-probability sequence of reactions $k_1, \dots, k_n$ that takes the state x to the state y .

²Such a function demonstrates that the dynamics has an attractive tendency towards a compact subset of the state-space.

this set-point encoding function cannot be arbitrary for maxRPA systems and must have a specific form depending on the stoichiometry of the first two reactions.

We shall allow disturbances to enter the system through perturbations of the propensity functions of all the reactions except the first two. These $(K - 2)$ reactions need not follow mass-action kinetics and we call the vector of their propensity functions $\lambda = (\lambda_3, \dots, \lambda_K)$ as the *uncertain propensity map* which belongs to some pre-defined set Λ_u of *uncertain* propensity maps. In our setup, constant-in-time disturbances affect the system by perturbing the uncertain propensity map λ to another map in the uncertain set Λ_u .

Given this framework, we now define the maxRPA property in a mathematically precise way. Let $\theta = (\theta_1, \theta_2)$ be the vector of rate constants for the first two reactions, which we assume belongs to some feasible open set $\Theta \subset \mathbb{R}_+^2$, and let $f : \mathbb{R}^N \rightarrow R$ be the projection map, that maps the state vector to the state of the output species

$$f(x_1, \dots, x_N) = x_1. \quad (2.9)$$

For any fixed $(\theta, \lambda) \in \Theta \times \Lambda_u$, let $(x_{\theta, \lambda}(t))_{t \geq 0}$ be the deterministic dynamics with fixed point $\bar{x}_{\theta, \lambda}$. Then we say that the network satisfies the maxRPA property in the *deterministic setting* if and only if

$$\lim_{t \rightarrow \infty} f(x_{\theta, \lambda}(t)) = f(\bar{x}_{\theta, \lambda}) = \phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right) \quad \text{for any} \quad (\theta, \lambda) \in \Theta \times \Lambda_u. \quad (2.10)$$

Similarly let $(X_{\theta, \lambda}(t))_{t \geq 0}$ be the stochastic dynamics with stationary distribution $\pi_{\theta, \lambda}$. Moreover let $\mathbb{E}$ (resp. $\mathbb{E}_\pi$) denote the expectation operator under the natural probability distribution (resp. the stationary probability distribution) for CTMC. Then we say that the network satisfies the maxRPA property in the *stochastic setting* if and only if

$$\lim_{t \rightarrow \infty} \mathbb{E}(f(X_{\theta, \lambda}(t))) = \mathbb{E}_{\pi_{\theta, \lambda}}(f) = \phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right) \quad \text{for any} \quad (\theta, \lambda) \in \Theta \times \Lambda_u. \quad (2.11)$$

2.3 RPA and Integrators

The *internal model principle* (IMP) [27, 29] from control theory implies that for a system to reject constant-in-time disturbances, it must contain a subsystem, called the *internal model* (IM), that is able to generate such disturbances. For this to hold, the subsystem must be an “integrator” as constant disturbances are generated by the ODE

$$\dot{z} = 0.$$

Additionally, IMP stipulates that the internal model IM is only affected by the system output, and it receives no other direct inputs - neither from the rest of the system, nor from the disturbance itself (see Figure 1). Intuitively, IM can be viewed as providing an estimate of the external disturbances, based *only* on the deviation of the output from its set-point.

Consider the deterministic setting and suppose there is a real-valued function $F(x)$ such that if we let $z_{\theta, \lambda}(t) = F(x_{\theta, \lambda}(t))$ then we have

$$\dot{z}_{\theta, \lambda}(t) = \phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right) - f(x_{\theta, \lambda}(t)). \quad (2.12)$$

In this case $z_{\theta,\lambda}(t)$ is an *integrator* whose time-derivative measures the deviation of the output from the set-point. The function $F(x)$ is encoded by the IM and it determines the change-of-coordinates that produces the integrator. Observe that for an integrator to achieve its objective of measuring output-deviations, we can allow for more flexibility in the r.h.s. of (2.12). For example, if $\Psi(x)$ is any nonlinear function of the state-vector $x = (x_1, \dots, x_N)$ which is monotonic in the first coordinate x_1 , then a more general version of the integrator equation (2.12) is given by

$$\dot{z}_{\theta,\lambda}(t) = \Psi \left(\phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right), y_{\theta,\lambda}(t) \right) - \Psi(x_{\theta,\lambda}(t)), \quad (2.13)$$

where $y_{\theta,\lambda}(t) = (x_{\theta,\lambda,2}(t), \dots, x_{\theta,\lambda,N}(t))$ denotes the last $(N-1)$ components of $x_{\theta,\lambda}(t)$. This integrator is able to sense the deviation of the output only if the dynamics of $y_{\theta,\lambda}(t)$ stays away those values y which make the mapping $x_1 \mapsto \Psi(x_1, y)$ identically zero. Typically $\Psi(x_1, y)$ is monomial of variables in y and in this case the integrator is often called *constrained* as the dynamics of $y_{\theta,\lambda}(t)$ is constrained to stay away from the boundary of $\mathbb{R}_+^{N-1}$ [37].

We now turn to the stochastic setting where the output exhibiting maxRPA is the expectation of the copy-number of species $\mathbf{X}_1$ (see (2.11)). In this case, for a real-valued function $F(x)$ to specify an integrator we must have that for $z_{\theta,\lambda}(t) = \mathbb{E}(F(X_{\theta,\lambda}(t)))$

$$\dot{z}_{\theta,\lambda}(t) = \phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right) - \mathbb{E}(f(X_{\theta,\lambda}(t))). \quad (2.14)$$

Under the assumption of ergodicity, such a function $F(x)$ exists uniquely (up to addition by a constant) and it can be found by solving the *Poisson* equation (see Chapter 8 in [38] and [39]). One key difference with the deterministic setting is that we cannot generalise the integrator equation through the introduction of an arbitrary nonlinear function Ψ (see (2.13)), as the expectation operator would not generally commute with the nonlinearity introduced by Ψ . As our results will indicate, this constraint in the form of the integrator equation manifests itself in structural constraints for the maxRPA networks.

We end this section with the simple observation that existence of an integrator is sufficient to guarantee maxRPA in both deterministic and stochastic setting, as the r.h.s. of the integrator equations need to be zero at steady-state.

3 Main Results

In this section we present the main results of our paper. These results provide the necessary and sufficient conditions for a network to achieve maxRPA in both deterministic and stochastic settings. The fundamental condition in both cases is a linear-algebraic relation for the stoichiometric structure of the network. Based on this relation we shall classify all maxRPA networks as either *homothetic* or *antithetic* and we shall show that in the stochastic scenario only the latter type of maxRPA networks can arise. We shall also explain how our results provide a novel *internal model principle* (IMP) for maxRPA networks.

Let S be the $N \times K$ stoichiometric matrix for the reaction network, defined as

$$S = \text{Col}(\zeta_1, \dots, \zeta_K), \quad (3.15)$$

where $\zeta_k = (\zeta_{1k}, \dots, \zeta_{dk})$ is the stoichiometry vector for reaction k . This matrix encodes the stoichiometric structure of the reaction network and we shall assume that the set of network reactions is *rich* enough to ensure full row-rank

$$\text{Rank}(S) = N. \quad (3.16)$$

Our principal maxRPA characterising linear-algebraic condition, in both deterministic and stochastic settings, is the existence of a vector $q = (q_1, \dots, q_N)$ and a positive constant κ satisfying

$$q^T S = (\kappa, -1, 0, \dots, 0). \quad (3.17)$$

Only the first two components of the r.h.s. vector are nonzero and they are κ and -1 . For stable networks, if such a pair (q, κ) exists, then it must be necessarily unique and it can be found by solving a simple linear system of the form $Ax = b$ (see Lemma S3.1 in the Supplement).

Under certain additional stoichiometric conditions, the structural condition of existence of (q, κ) satisfying (3.17) characterises all maxRPA systems in both deterministic and stochastic settings. These additional conditions are stricter in the stochastic case, which shows that the class of maxRPA systems in the stochastic scenario is a strict subset of the class of such systems in the deterministic scenario. In the next couple of sections we explicitly state our assumptions and our characterisation results in both deterministic and stochastic settings.

3.1 Deterministic maxRPA networks

We now characterise maxRPA networks in the deterministic setting and explicitly identify the form of the set-point encoding function (2.8) as well as an *integrator* that demonstrates the maxRPA property (see Section 2.3). Recall that in our setting, the first two reactions have mass-action propensities with rate constants $\theta = (\theta_1, \theta_2) \in \Theta$. Hence these propensity functions can be expressed as $\lambda_k(x) = \theta_k m_k(x)$ for $k = 1, 2$, where m_k denotes the mass-action monomial

$$m_k(x) = \prod_{i=1}^N x_i^{\nu_{ik}}. \quad (3.18)$$

Disturbances enter the reaction system via uncertainties or perturbations in the propensities of the other $K - 2$ reactions. These $K - 2$ propensities constitute the uncertain propensity map λ which belongs to the uncertain set Λ_u . The deterministic reaction dynamics $(x_{\theta, \lambda}(t))_{t \geq 0}$ satisfies the following system of ODEs

$$\frac{dx}{dt} = \sum_{k=1}^2 \theta_k m_k(x) \zeta_k + \sum_{k=3}^K \lambda_k(x) \zeta_k. \quad (3.19)$$

We shall make the following assumptions.

Assumption 3.1 *The full row-rank condition (3.16) holds and for any fixed $(\theta, \lambda) \in \Theta \times \Lambda_u$ we have the following:*

- (A) **Stability:** The ODE system (3.19) is well-defined for initial values in some open set $\mathcal{E} \subset \mathbb{R}_+^N$ which is time-invariant under the dynamics and there is a globally attracting fixed point $\bar{x}_{\theta,\lambda} \in \mathcal{E}$.
- (B) **Local richness of the uncertain set Λ_u :** There exist R functions $\phi_1, \dots, \phi_R : \mathcal{E} \rightarrow \mathbb{R}_+^{K-2}$ such that for some $\epsilon > 0$

$$\lambda_\gamma := \lambda + \sum_{i=1}^R \gamma_i \phi_i \in \Lambda_u \quad \text{for all } \gamma = (\gamma_1, \dots, \gamma_R) \in (-\epsilon, \epsilon)^R \quad (3.20)$$

and if we define a $(K-2) \times R$ matrix as

$$\Phi(x) = \text{Col}(\phi_1(x), \dots, \phi_R(x)) \quad (3.21)$$

then this matrix has full row-rank at $x = \bar{x}_{\theta,\lambda}$.

Part (B) of this assumption implies that around any uncertain propensity map λ , the uncertain set Λ_u is *locally rich* in the sense that it is possible to solely perturb the propensity for each disturbance inducing reaction $k \in \{3, \dots, K\}$ by translating the propensity map λ by a linear combination of functions $\phi_1, \dots, \phi_R$. Typically we would expect that each propensity gets independently disturbed, i.e. $R = K - 2$ and $\phi_k(x) = (0, \dots, \hat{\phi}_k(x), 0, \dots, 0)$. In this case, the rank condition is trivially satisfied and the condition in part (B) reduces to saying that for each $k = 1, \dots, K - 2$, the propensity $\lambda_{k+2}(x)$ can be perturbed to

$$\lambda_{k+2}(x) + \gamma \hat{\phi}_k(x) \quad \text{for all } \gamma \in (-\epsilon, \epsilon).$$

In the special case of mass-action kinetics, i.e.

$$\lambda_{k+2}(x) = \theta_{k+2} m_{k+2}(x),$$

we have

$$\hat{\phi}_k(x) = m_{k+2}(x),$$

and this condition implies that the disturbance perturbs the rate constant θ_k to a value in $(\theta_k - \epsilon, \theta_k + \epsilon)$. While this covers many scenarios of interest, the general condition, as stated in Assumption 3.1(B), is satisfied by many other forms of disturbances, like those arising from dependent sources like metabolic burden, cross-talk, temperature and other extrinsic factors etc.

Observe that if some component of the uncertain propensity map λ is the zero function, then one cannot perturb it negatively, and hence (3.20) will not hold for all $\gamma \in (-\epsilon, \epsilon)^R$. However our characterisation result will remain valid as its proof will continue to hold upon replacing the usual derivatives w.r.t. components of γ with appropriate one-sided derivatives. A zero component in the uncertain propensity map λ will naturally arise when one wants to check if adding a disturbance inducing reaction would preserve or violate the maxRPA property.

Theorem 3.2 (Characterisation of deterministic maxRPA networks) Suppose that the network reaction dynamics is given by (3.19) and Assumption 3.1 holds. Then this network exhibits maxRPA for the output species $\mathbf{X}_1$ if and only if the following two stoichiometric conditions are satisfied:

(A) Reactions 1 and 2 have as reactants, strictly unequal number of molecules of the output species $\mathbf{X}_1$ but equal number of molecules of all other network species, i.e.

$$\nu_{11} \neq \nu_{12} \quad \text{but} \quad \nu_{i2} = \nu_{i1} \quad \text{for each} \quad i = 2, \dots, N. \quad (3.22)$$

(B) There exists a pair (q, κ) , with $\kappa > 0$, satisfying the linear system (3.17).

Moreover, if the network exhibits maxRPA, the set-point encoding function is given by

$$\phi_{\text{out}}(x) = (\kappa x)^{\frac{1}{\nu_{12} - \nu_{11}}} \quad (3.23)$$

and an integrator is given by the function

$$F(x) := \frac{1}{\theta_2} \sum_{i=1}^N q_i \int \frac{1}{m_1(x)} dx_i = \frac{1}{\theta_2} \sum_{i=1}^N \frac{q_i v_i(x_i)}{\prod_{j=1, j \neq i}^N x_j^{\nu_{j1}}} \quad (3.24)$$

where

$$v_i(x) = \int \frac{1}{x^{\nu_{i1}}} dx = \begin{cases} \log x & \text{if } \nu_{i1} = 1 \\ \frac{1}{1 - \nu_{i1}} x^{1 - \nu_{i1}} & \text{otherwise.} \end{cases}$$

Remark 3.3 Observe that the gradient of integrator (3.24) is given by

$$\nabla_x F(x) = \frac{1}{\theta_2 m_1(x)} q$$

which shows that if q satisfies the linear system (3.17) then for $z_{\theta, \lambda}(t) = F(x_{\theta, \lambda}(t))$ we have

$$\dot{z}_{\theta, \lambda}(t) = \kappa \frac{\theta_1}{\theta_2} - \frac{m_2(x_{\theta, \lambda}(t))}{m_1(x_{\theta, \lambda}(t))} = \kappa \frac{\theta_1}{\theta_2} - (f(x_{\theta, \lambda}(t)))^\alpha, \quad (3.25)$$

where $\alpha := \nu_{12} - \nu_{11}$ and $f(x) = x_1$. The last relation holds because $m_2(x) = x_1^\alpha m_1(x)$ due to condition (3.22). Instead of (3.24) we can also define a simpler linear integrator

$$F_c(x) = \frac{1}{\theta_2} \sum_{i=1}^N q_i x_i.$$

In this case for $z_{\theta, \lambda}(t) = F_c(x_{\theta, \lambda}(t))$ we have

$$\dot{z}_{\theta, \lambda}(t) = m_1(x_{\theta, \lambda}(t)) \left(\kappa \frac{\theta_1}{\theta_2} - (f(x_{\theta, \lambda}(t)))^\alpha \right). \quad (3.26)$$

Both (3.25) and (3.26) are examples of the general integrator equation (2.13).

3.2 Stochastic maxRPA networks

In this section we provide a result analogous to Theorem 3.2 for maxRPA networks in the stochastic setting. We shall need to make assumptions similar to Assumption 3.1 but these assumptions are more technical in nature in the stochastic setting. We now informally describe these assumptions and their formal description can be found in Section S3.2 of the Supplement.

Assumption 3.4 For any $(\theta, \lambda) \in \Theta \times \Lambda_u$ we have the following:

- (A) **Stability:** The state-space $\mathcal{E} = \mathbb{N}_0^N$ is irreducible for the CTMC dynamics $(X_{\theta, \lambda}(t))_{t \geq 0}$ and this CTMC is exponentially ergodic with a unique stationary probability distribution $\pi_{\theta, \lambda}$.
- (B) **Richness of the uncertain set Λ_u :** This criterion is based on the solution of the Poisson equation that defines the integrator (2.14). It essentially implies that the set Λ_u of uncertain propensity maps is sufficiently rich to allow perturbations by functions whose linear combination can closely approximate the Poisson equation solution in a certain norm defined using the stationary probability distribution $\pi_{\theta, \lambda}$.

The full-rank condition (3.16) is necessary for the full nonnegative integer orthant $\mathbb{N}_0^N$ to be an irreducible state-space for the stochastic dynamics (e.g. see [35]). In fact for most biological reaction networks of interest, this condition is also sufficient in ensuring that $\mathbb{N}_0^N$ is irreducible. Moreover it has been found that many such networks are indeed exponentially ergodic [36] and this can be shown by constructing a linear Foster-Lyapunov function. The condition on the richness of the uncertain set Λ_u is more demanding than part (B) of Assumption 3.1, to account for the fact that even at stationarity the dynamics is fluctuating over the whole state-space $\mathcal{E}$.

Theorem 3.5 (Characterisation of stochastic maxRPA networks) Suppose that the network reaction dynamics is given by the CTMC $(X_{\theta, \lambda}(t))_{t \geq 0}$ and also suppose that Assumption 3.4 holds. Then this network exhibits maxRPA for the output species $\mathbf{X}_1$ if and only if the following two stoichiometric conditions are satisfied:

- (A) Reactions 1 and 2 **do not have** molecules of species $\mathbf{X}_2, \dots, \mathbf{X}_N$ as reactants, and **exactly one** of them involves a **single** molecule of the output species $\mathbf{X}_1$ as reactant while the other involves **none** i.e.

$$\begin{aligned} \nu_{i2} = \nu_{i1} = 0 \quad \text{for each } i = 2, \dots, N \\ \text{and either } \nu_{12} = 1 - \nu_{11} = 1 \quad \text{or} \quad \nu_{11} = 1 - \nu_{12} = 1. \end{aligned} \quad (3.27)$$

- (B) There exists a pair (q, κ) , with $\kappa > 0$, satisfying the linear system (3.17).

Moreover, if the network exhibits maxRPA, the set-point encoding function is given by

$$\phi_{\text{out}}(x) = (\kappa x)^{\frac{1}{\nu_{12} - \nu_{11}}} \quad (3.28)$$

and an integrator is given by the linear function

$$F(x) = \frac{1}{\theta_2} \sum_{i=1}^N q_i x_i. \quad (3.29)$$

Remark 3.6 Observe that the stoichiometric conditions imposed by this theorem are stricter than those imposed by Theorem 3.2. This shows that if the underlying assumptions hold, then the set of networks exhibiting maxRPA in the stochastic setting is a strict subset of the set of networks exhibiting maxRPA in the deterministic setting. Also note that in this stochastic case the difference in the number of reactant molecules for the output species $\mathbf{X}_1$, between reactions 1 and 2 is either 1 or -1 . Hence $\nu_{12} - \nu_{11} \in \{-1, 1\}$ and so the set-point encoding function can also be expressed as

$$\phi_{\text{out}}(x) = (\kappa x)^{\nu_{12} - \nu_{11}}.$$

3.3 Homothetic and Antithetic maxRPA networks

In the previous sections, we characterised networks achieving maxRPA in both deterministic and stochastic settings. These networks comprise N species $\mathbf{X}_1, \dots, \mathbf{X}_N$ and from now on we add the null species $\mathbf{X}_0 = \emptyset$ in our species-set and we adopt the convention that

$$\mathbf{X}_i + \mathbf{X}_0 \equiv \mathbf{X}_i$$

for any species $\mathbf{X}_i$. Our key stoichiometric criterion for RPA characterisation, which is common to both deterministic and stochastic settings, is the existence of a unique solution (q, κ) to the linear-algebraic system (3.17). Letting $q = (q_1, \dots, q_N)$, we can think of q_i as the *charge* of species $\mathbf{X}_i$ and to the null species $\mathbf{X}_0$ we assign the charge of $q_0 = 0$. Based on these charge values, we can partition the species set $\mathcal{C} = \{\mathbf{X}_0, \dots, \mathbf{X}_N\}$ of a maxRPA network into the set of positively charged species $\mathcal{C}_+$, negatively charged species $\mathcal{C}_-$ and neutral species $\mathcal{C}_0$

$$\mathcal{C}_+ = \{\mathbf{X}_i \in \mathcal{C} : q_i > 0\}, \quad \mathcal{C}_- = \{\mathbf{X}_i \in \mathcal{C} : q_i < 0\} \quad \text{and} \quad \mathcal{C}_0 = \{\mathbf{X}_i \in \mathcal{C} : q_i = 0\}.$$

Based on this partition we classify all maxRPA networks as follows:

Definition 3.7 (Classification of maxRPA networks) A maxRPA network is called **homothetic** if either $\mathcal{C}_+$ or $\mathcal{C}_-$ is empty. On the other hand, a maxRPA network is called **antithetic** if both $\mathcal{C}_+$ and $\mathcal{C}_-$ are nonempty.

In other words, for a maxRPA network to be homothetic, all the nonzero entries in vector q must have the same sign, while if q has least two nonzero entries with opposite signs, then this network is antithetic.

The linear-algebraic condition (3.17) implies that for each disturbance inducing reaction $k \in \{3, \dots, K\}$, its stoichiometric vector ζ_k is orthogonal to q , i.e. $q^T \zeta_k = 0$. Hence each of these reactions are *charge-preserving* in the sense that the cumulative charge of the reactant complex is the same as that of the product complex

$$\sum_{i=1}^N q_i \nu_{ik} = \sum_{i=1}^N q_i \nu'_{ik}.$$

Consider the common scenario where all reactions are *bimolecular*, i.e. all reactions can have at most two reactants i.e. $\sum_i \nu_{ik} \leq 2$. In this scenario antithetic RPA networks are marked by the presence of a generalised *sequestration* or *annihilation* reaction of the form

where $\mathbf{X}_p \in \mathcal{C}_+$, $\mathbf{X}_\ell \in \mathcal{C}_-$ and $\mathcal{C}_0^*$ denotes any combination of species that belong to the set $\mathcal{C}_0$ (see Remark S3.5 in the Supplement). Note that such a reaction cannot arise in a homothetic network because either $\mathcal{C}_+$ or $\mathcal{C}_-$ is empty.

3.4 A novel *Internal model principle* (IMP)

We now discuss how our characterisation results provide a novel IMP for maxRPA networks, as illustrated in Figure 1. Let $\mathcal{C}_\pm := \mathcal{C}_+ \cup \mathcal{C}_-$ denote the set of species that form the support for vector q i.e. $q_i \neq 0$ if and only if $\mathbf{X}_i \in \mathcal{C}_\pm$. We can view the species in $\mathcal{C}_\pm$ as forming the *internal model* (IM) that generates external disturbances via the integrator (see Section 2.3) defined by function $F(x)$ identified in Theorems 3.2 and 3.5.

Typically, one is interested in situations where the disturbance inducing reactions are able to independently perturb the output species $\mathbf{X}_1$

$$e_1 \in \text{Range}(\zeta_3, \dots, \zeta_K). \quad (3.31)$$

In particular, if there exists a reaction that creates or degrades a molecule of $\mathbf{X}_1$, then this condition is trivially satisfied. An example where this condition does not hold is the simple birth-death module

which may include several disturbance inducing reactions that do not alter the copy-numbers of the output species $\mathbf{X}_1$. In this example the IM includes the output species $\mathbf{X}_1$ and the property of RPA is trivial because the output is unaffected by the disturbances. On the other hand, if condition (3.31) holds then the IM cannot contain the output species $\mathbf{X}_1$ (see Lemma S3.2 in the Supplement) and for it to perform integral action and generate robustness, there must exist one or more actuation reactions from the IM to the rest of the network (RON). These actuation reactions complete the feedback loop and their presence is necessary for the stability of the full RPA network (see Figure 1). Generally actuation reactions are a subset of disturbance inducing reactions, but it is possible for reaction 1 and/or 2 to serve as actuation reactions (see Section 5.1).

3.5 All stochastic maxRPA networks are Antithetic

Our next result shows that in the stochastic setting any maxRPA network must embed the so-called *antithetic* motif, which was first discovered in [8], thereby showing that the antithetic topology is “universal” in the stochastic setting. This result extends a similar universality result proved in [24], for the situation inspired by synthetic biology applications, where the controller network is separate from the network being controlled, and the latter is arbitrary. We shall assume that the network is bimolecular, and as mentioned in Section 3.3 a key property of antithetic networks is the existence of a generalised annihilation or sequestration reaction of the form (3.30) between the IM (comprising species in $\mathcal{C}_\pm$) and the RON (comprising species in $\mathcal{C}_0$).

Theorem 3.8 (Generic decomposition of stochastic maxRPA networks) *Suppose that Assumption 3.4 and condition (3.31) holds and all reactions are bimolecular. Then*

Figure 2: **Generic decomposition of stochastic maxRPA networks.** Depicts the generic structure of a maxRPA network in the stochastic setting, as proved by Theorem 3.8. Such a network decomposes into an *internal model* (IM), comprising species in $\mathcal{C}_{\pm} = \mathcal{C}_{+} \cup \mathcal{C}_{-}$, and the *rest of the network* (RON) with species in $\mathcal{C}_0$. There is a set-point encoding reaction (3.32) producing species $\mathbf{X}_2 \in \mathcal{C}_{+}$ and a sensing reaction (3.33) that measures the output species $\mathbf{X}_1$ via production of species $\mathbf{X}_3 \in \mathcal{C}_{-}$. All reactions with reactants in IM and products in RON must have the generalised annihilation form (3.30) and there must exist at least one such reaction, as indicated.

for a maxRPA network in the stochastic setting we have the following up to renaming of species $\mathbf{X}_2, \dots, \mathbf{X}_N$ and interchanging of reactions 1 and 2:

(A) The maxRPA network must be antithetic.

(B) There is a species $\mathbf{X}_2 \in \mathcal{C}_{+}$ such that the first reaction is of the form

where $\mathcal{C}^*$ denotes any combination of network species.

(C) There is a species $\mathbf{X}_3 \in \mathcal{C}_{-}$, such that the second reaction is of the form

(D) There exists at least one generalised sequestration reaction of the form (3.30). Moreover reactions of this form are the only possible reactions where the products are species in $\mathcal{C}_0$ and one of the reactant species is in $\mathcal{C}_{\pm}$.

Remark 3.9 We refer to (3.32) as the set-point encoding reaction (as its reaction rate θ_1 sets the numerator of the set-point) and to (3.33) as the output sensing reaction (as its reaction rate is $\theta_2 x_1$ and x_1 is the output species copy-number). The encoding and sensing is performed by the internal model (IM) (comprising species in $\mathcal{C}_{\pm}$) via production of distinct species $\mathbf{X}_2$ and $\mathbf{X}_3$ (the species are renamed if necessary) in the IM. The generic decomposition of stochastic maxRPA networks is depicted in Figure 2.

4 maxRPA networks with minimal IMs

The internal model (IM) of a maxRPA network characterises the embedded control mechanism that generates the RPA property. In this section we identify the minimal IM necessary for maxRPA where by minimality we mean the IM with the least number of species and the least number of reactions, in that order. This minimality question has important consequences from both systems and synthetic biology perspectives. For example, one might surmise that evolution will seek out the least complex designs for the IM module as they would bestow maxRPA with the smallest amount of extra metabolic burden [21, 22]. This also makes minimal complexity an especially useful design criterion in synthetic biology where building complex circuits is fraught with many difficulties like loading effects [40], cross-talk [20] etc.

Of course the least complex IM would be the one which has only one species. We call this a *singleton* IM and in Section 4.1 we explore the structure of maxRPA networks with such IMs. Observe that these networks must be *homothetic* and the dynamics must be deterministic as in the stochastic setting any maxRPA network must necessarily be antithetic and its IM should comprise least two species (see Theorem 3.8). Hence networks with singleton IM only exhibit maxRPA in the deterministic setting and not the stochastic setting, illustrating that the containment of the set of stochastic maxRPA networks within the set of deterministic maxRPA networks is strict (see Remark 3.6). The least complex IM for an antithetic maxRPA network would consist of exactly two species and we refer to such an IM as *dyadic*. In Section 4.2 we explore the structure of maxRPA networks with dyadic IMs.

4.1 Homothetic networks with *singleton* IMs

Consider a deterministic maxRPA network whose IM consists of only the species $\mathbf{X}_2$, i.e. $\mathcal{C}_\pm = \{\mathbf{X}_2\}$. By definition this network is homothetic and we shall assume that condition (3.31) holds. Let (q, κ) be the maxRPA characterising pair that satisfies (3.17). Then only the second component of q (i.e. q_2) is non-zero while the rest are all zeros. From (3.17) we can conclude that the copy-numbers of $\mathbf{X}_2$ are not altered by the disturbance inducing reactions (i.e. $\zeta_{2k} = 0$ for $k = 3, \dots, K$) and for the first two reactions we have

$$n_1 := \zeta_{21} = \frac{\kappa}{q_2} \quad \text{and} \quad n_2 := -\zeta_{22} = \frac{1}{q_2}.$$

Recall that the nonnegative integer ν_{ik} denotes the number of molecules of species $\mathbf{X}_i$ that are consumed as reactants by the k -th reaction. Due to Theorem 3.2 the first two reactions, with mass-action kinetics, must have the form

where α is some non-zero integer satisfying $\alpha \geq -\nu_{11}$ and $\mathcal{C}_0^*$ denotes any combination of species in $\mathcal{C}_0 = \{\mathbf{X}_1, \mathbf{X}_3, \dots, \mathbf{X}_N\}$. Recall from Remark 3.9, that reaction 1 is the

Figure 3: **Deterministic maxRPA networks with singleton internal models (homothetic) and zeroth-order degradation.** The internal model (IM) consists of only one species $\mathcal{C}_{\pm} = \{\mathbf{X}_2\}$ whose charge q_2 is either positive (panel A) or negative (panel B). All the other species belong to the rest of the network (RON). The zeroth-order degradation reaction is shown with a dotted arrow.

set-point encoding reaction, while reaction 2 is the output sensing reaction. Note that since $\kappa > 0$, both n_1 and n_2 must have the same sign as q_2 and the output set-point encoding is given by

$$\theta^* = \phi_{\text{out}} \left(\frac{\theta_1}{\theta_2} \right) = \sqrt[\alpha]{\frac{\theta_1 n_1}{\theta_2 n_2}}.$$

4.1.1 Zeroth-order degradation networks

Consider the case $\nu_{21} = 0$ in which the lone IM species $\mathbf{X}_2$ is not a reactant in either the first reaction (4.34) or the second reaction (4.35). This implies that exactly one of these two reactions causes *negative production* or loss of $\mathbf{X}_2$ molecules, violating the reaction form (2.1). The rate of this reaction is not proportional to the abundance of $\mathbf{X}_2$ molecules which can lead to negative abundance levels for this species. Such a reaction is called *zeroth-order degradation* and even though it is non-physical it can be approximately realised when the degradation is assisted by an enzyme which operates at saturation [10].

Depending on whether q_2 is positive (i.e. $\mathcal{C}_+ = \{\mathbf{X}_2\}$) or negative (i.e. $\mathcal{C}_- = \{\mathbf{X}_2\}$), two types of maxRPA networks based on zeroth-order degradation can be constructed. An example of both these network types is shown in Figure 3. In both the examples, we set $\nu_{i1} = 0$ for all i . The key difference between them is that q_2 is positive in one case (panel A) while it is negative in the other (panel B). In the former, the zeroth-order degradation reaction acts like a sensing reaction while in the latter, it acts like a set-point encoding reaction.

4.1.2 Autocatalytic networks

Now consider the case $\nu_{21} > 0$ in which the lone IM species $\mathbf{X}_2$ is a reactant in both the first reaction (4.34) and the second reaction (4.35). Hence, irrespective of the sign of q_2 , $\mathbf{X}_2$ stimulates its own production and therefore we call such networks *autocatalytic*.

In Figure 4 we present two autocatalytic maxRPA networks that resemble the networks in Figure 3, except that non-physical zeroth-order degradation reactions are replaced by autocatalytic reactions. In both the examples we set $\nu_{i1} = 0$ for all i except $i = 2$. As before, the difference between them is that q_2 is positive in one case

Figure 4: **Deterministic maxRPA networks with singleton internal models (homothetic) and autocatalytic reactions.** The internal model (IM) consists of only one species $\mathcal{C}_{\pm} = \{\mathbf{X}_2\}$ whose charge q_2 is either positive (panel A) or negative (panel B). All the other species belong to the rest of the network (RON). The autocatalytic reaction is shown by a loop.

(panel A) while it is negative in the other (panel B). In the former, we choose $\nu_{21} = n_2$ and so reaction 1 (set-point encoding) is autocatalytic, while in the latter we choose $\nu_{21} = -n_1$ and so reaction 2 (output sensing) is autocatalytic.

In the special case of $\nu_{21} = -n_1 = n_2 = 1$, $\mathcal{C}_0^* = \mathbf{X}_1$ in (4.34) and $\mathcal{C}_0^* = \mathbf{X}_0 = \emptyset$ in (4.35), these two reactions encode the simplest network exhibiting the property of *Absolute Concentration Robustness* (ACR), whereby one of the species ($\mathbf{X}_1$ in our case) has the same value at any equilibrium value. For more details we refer the readers to the celebrated paper of Shinar and Feinberg [41] which present sufficient structural conditions for networks to exhibit ACR. In a recent paper the existence of a linear *constrained* integrator for a wide class of ACR networks is shown [42].

4.2 Antithetic maxRPA networks with *dyadic* IMs

Consider an antithetic maxRPA network in the deterministic setting whose IM consists of two species $\mathbf{X}_2$ and $\mathbf{X}_3$, i.e. $\mathcal{C}_{\pm} = \{\mathbf{X}_2, \mathbf{X}_3\}$. Let (q, κ) be the maxRPA characterising pair that satisfies (3.17). Without loss of generality we can assume that $q_2 > 0$ (i.e. $\mathcal{C}_+ = \{\mathbf{X}_2\}$) and $q_3 < 0$ (i.e. $\mathcal{C}_- = \{\mathbf{X}_3\}$). All the other components of the q vector are zero. In this section we shall assume that all reactions are *bimolecular* i.e. $\sum_i \nu_{ik} \leq 2$ for each reaction k of the form (2.1).

From (3.17) we can conclude that for each disturbance inducing reaction $k = 3, \dots, K$ we must have

$$q_2 \zeta_{2k} = -q_3 \zeta_{3k}, \quad (4.36)$$

and so ζ_{2k} and ζ_{3k} have the same sign. Hence each disturbance inducing reaction either simultaneously *degrades* or *produces* both species $\mathbf{X}_2$ and $\mathbf{X}_3$. Suppose that one such reaction k causes degradation of both the species, then as this reaction is bimolecular it must be of the generalised sequestration (3.30) form

Since $\zeta_{2k} = \zeta_{3k} = -1$, due to (4.36) we have $q_3 = -q_2$. To preserve dynamical stability, such a degradation causing reaction must indeed exist provided one of species $\mathbf{X}_2$ or $\mathbf{X}_3$ is only produced by the first two reactions, and not degraded.

Let us examine the form of the first two reactions more closely under the restrictions imposed by Theorem 3.2. Without loss of generality we can assume that $\alpha > 0$, as the other case is symmetric. As all reactions are bimolecular, there must exist a species $\mathbf{X}_m$ such that the first two reactions, with mass-action kinetics, have the form

Here $\mathbf{X}_m$ can be any species, including the null species $\mathbf{X}_0 = \emptyset$ for $m = 0$. Note that for the second reaction to be bimolecular, α can be 2 only when $m = 0$, and for other values of m , α should be 1. We now separate our analysis into two cases, depending on the value of m .

Case 1: $m \notin \{2, 3\}$: In this case $q_m = 0$ and (3.17) implies that

$$\nu'_{21}q_2 + \nu'_{31}q_3 = \kappa > 0. \quad \text{and} \quad \nu'_{22}q_2 + \nu'_{32}q_3 = -1.$$

Since $q_2 > 0$ and $q_3 < 0$ we must have that $\nu'_{21} > 0$ and $\nu'_{32} > 0$ which says that the first reaction must produce species $\mathbf{X}_2$ while the second reaction must produce species $\mathbf{X}_3$. As these two reactions are not degrading any $\mathbf{X}_2$ or $\mathbf{X}_3$, there must exist a disturbance inducing reaction of the form (4.37), as we explained earlier, which in turn implies that $q_3 = -q_2$. We illustrate this case with an example shown in Figure 5 where we set $q_2 = -q_3 = \nu'_{21} = \nu'_{32} = \alpha = 1$ and $\nu'_{31} = \nu'_{22} = 0$. Note that κ must be 1 for this example and so the set-point is exactly the ratio θ_1/θ_2 . If $m = 0$ this antithetic network is also maxRPA in the stochastic setting (see Theorem 3.5).

Case 2: $m \in \{2, 3\}$: Without loss of generality we assume that $m = 2$ as the other case ($m = 3$) is similar. Then (3.17) implies that

$$(\nu'_{21} - 1)q_2 + \nu'_{31}q_3 = \kappa > 0. \quad \text{and} \quad (\nu'_{22} - 1)q_2 + \nu'_{32}q_3 = -1.$$

As $q_2 > 0$ and $q_3 < 0$, the first reaction must produce at least two molecules of $\mathbf{X}_2$ (i.e. $\nu'_{21} \geq 2$). If either ν'_{31} or ν'_{32} is positive then species $\mathbf{X}_3$ is being produced by these reactions but not being degraded, and so $q_2 = -q_3$ and the two IM species must degrade mutually via a generalised annihilation reaction (4.37), giving rise to a similar antithetic structure as in the previous case. On the other hand if $\nu'_{31} = \nu'_{32} = 0$, then species $\mathbf{X}_3$ is not getting created by the first two reactions, and we must necessarily have $\nu'_{22} = 0$, implying that reaction 2 causes degradation of $\mathbf{X}_2$ molecules, without the need of additional annihilation reactions. In this scenario, species $\mathbf{X}_3$ merely acts as a bystander and does not participate in creating the integral action that generates robustness. This integral action is being created only by species $\mathbf{X}_2$ as in the case of homothetic autocatalytic networks with a singleton internal model (see Section 4.1.2).

5 Biological examples

The aim of this section is to illustrate our results with known examples of maxRPA networks from existing literature. In the context of synthetic biology, antithetic maxRPA networks can be engineered by interfacing any intracellular network with the antithetic

Figure 5: **A deterministic antithetic maxRPA network.** The internal model (IM) has only two species $\mathcal{C}_{\pm} = \{X_2, X_3\}$ with $q_2 = -q_3 > 0$. A species X_m (with $m \neq 0$) belonging to the rest of the network (RON) catalyses the production of these two species via reactions 1 and 2. These two species degrade mutually via a generalised annihilation reaction (4.37). This network is also maxRPA in the stochastic setting if and only if $m = 0$.

integral feedback (AIF) controller [8]. This was successfully realised in bacteria [24], in vitro [43] and recently in mammalian cells [44]. For *E. Coli*, a slight variant of this approach was adopted in the design of the quasi-integral controller [32] and it is speculated that the AIF controller might be involved in the regulation of housekeeping genes [8]. While it is an ongoing effort to find more examples of endogenous networks that fit the antithetic paradigm, we focus on deterministic homothetic maxRPA networks in the rest of this section. Unless otherwise stated, all reactions follow mass-action kinetics in the upcoming examples.

5.1 Bacterial chemotaxis network

Chemotaxis refers to the movement of cells towards chemo-attractants or away from chemo-repellants [4, 30]. In bacterial cells such as *E. Coli* the mechanism of chemotaxis is well-understood and it is characterised by alternating phases of random tumbling and straight-line motion. Upon addition of an attractant the tumbling frequency first decreases sharply but then adapts perfectly to the previous level, demonstrating the phenomenon of RPA. In fact for a simplified model of bacterial chemotaxis (see Figure 6), adapted from [45], this RPA is maximal and we illustrate this using our characterisation result (Theorem 3.2).

The networks consists of three species - the active methylated receptor X_1 , the inactive methylated receptor X_2 and the unmethylated receptor X . These species interact through four reactions:

Figure 6: Bacterial chemotaxis network. This figure depicts a simplified model of bacterial chemotaxis adapted from [45]. The two key players are the active and the inactive forms of the methylated receptor, denoted by $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively. In the latter an attractor molecule is bound to the receptor while in the former the receptor is free. The two forms reversibly switch between each other and the rate of switching from $\mathbf{X}_1$ to $\mathbf{X}_2$ depends on the attractor concentration a . In the active form, the methylated receptor $\mathbf{X}_1$ can get demethylated by the phosphatase CheB to produce the deactivated receptor $\mathbf{X}$, which in turn can get methylated by enzyme CheR to produce $\mathbf{X}_1$. The RPA output is the concentration of $\mathbf{X}_1$ which determines the bacterial tumbling rate.

where θ_1 and θ_2 are determined by the concentrations of enzymes CheR and CheB respectively, and a denotes the concentration of the attractor. Assuming the unmethylated receptors are in abundance, we can replace $\mathbf{X}$ by the null species $\mathbf{X}_0 \equiv \emptyset$ in these reactions to obtain the following 2×4 stoichiometry matrix

$$S = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Notice that condition (3.22) holds, and for $q = (1, 1)$ and $\kappa = 1$ the linear system (3.17) is also satisfied. Therefore by Theorem 3.2, this network exhibits maxRPA with output set-point θ_1/θ_2 and an integrator is given by the function

$$F(x_1, x_2) = \frac{x_1 + x_2}{\theta_2}.$$

In this example, methylation is the *set-point encoding* reaction, demethylation is the *sensing* or the *measurement* reaction, while the attractor binding and unbinding reactions are the *disturbance inducing* reactions.

Observe that if $\mathbf{X}$ is not the null species $\emptyset$, then instead of a zero-order process, the methylation of the unmethylated receptor by CheR becomes a first-order process. In

Figure 7: **Proliferative network for cell-lineages.** This figure depicts a motif adapted from [31] for control of proliferative cell lineages. This N -species network begins with a pluripotent stem cell $\mathbf{X}_N$, progresses through $(N - 2)$ intermediate progenitor stages $\mathbf{X}_{N-1}, \dots, \mathbf{X}_2$ and then ends with a terminal differentiated cell $\mathbf{X}_1$. At each stage (except the last) the cells can *self-replicate* or progress to the next stage. The terminal cell can spontaneously degrade and it can also inhibit the proliferative tendency of the stem cell $\mathbf{X}_N$. The RPA output is the concentration of the terminal cell $\mathbf{X}_1$ which determines the size of the tissue in organ development and influences its maintenance and regeneration.

this case, condition (3.22) will get violated and the network will not exhibit maxRPA. This loss of maxRPA also occurs if CheB is allowed to demethylate both active and inactive receptors ($\mathbf{X}_1$ and $\mathbf{X}_2$), as we cannot find a solution to the linear system (3.17) in this case. These observations regarding network constraints for the RPA property have already been made for this chemotaxis network (see [2]), but our characterisation result facilitates drawing such conclusions for arbitrarily complex networks.

5.2 Proliferative control of cell lineages

Cell lineages are basic units of tissue and organ development, and the process of lineage proliferation is characterised by several stages of distinct progenitor cells that culminate into mature, highly-differentiated and non-dividing terminal cells that realise various tissue-functions [46]. This process is initiated by pluripotent stem cells and at each stage the progenitor cells can either self-replicate or progress to the next stage, with probabilities depending on the expression of certain stage-specific genes [47]. It is well-known that this lineage proliferation process is tightly-regulated and highly-robust, in the sense that the concentration of terminal cells is insensitive to the stage-specific probabilities that decide if a progenitor cell self-renews or passes to the next stage. This robustness is crucial for development and also for proper tissue repair following an injury.

It was shown computationally and experimentally in [31] that this robustness is achieved when the terminal cells secrete a chemical that can inhibit the self-renewal of a progenitor cell. For maximal robustness, this progenitor cell must be the stem

cell and we explore the property of maxRPA by considering an adapted version of the cell-lineage proliferation network given in [31] (see Figure 7). This network consists of N species given by $\mathbf{X}_1, \dots, \mathbf{X}_N$, with $\mathbf{X}_N$ being the stem cell, $\mathbf{X}_1$ being the terminal differentiated cell, and $\mathbf{X}_2, \dots, \mathbf{X}_{N-1}$ being the progenitor cells at $(N-2)$ intermediate stages. These species interact through the following $(2N-1)$ reactions:

Observe that for any intermediate progenitor cell $\mathbf{X}_i$ the chance that it will self-replicate, rather than progress to the next stage, is given by the ratio $r_i/(r_i + k_i)$. For the stem cell this probability becomes

$$\frac{\theta_1}{\theta_1 + \theta_2 x_1},$$

where x_1 is the terminal cell concentration. This clearly shows the negative feedback exerted by the terminal cell on the stem cell by suppressing its tendency to self-replicate.

One can see that condition (3.22) holds and by writing the $N \times (2N-1)$ stoichiometry matrix one can easily check that for $q = (1, 0, \dots, 0)$ and $\kappa = 1$ the linear system (3.17) is satisfied. Hence by Theorem 3.2 we can conclude that this network shows maxRPA with output set-point θ_1/θ_2 . Therefore the concentration of the terminal cells is robustly maintained at the ratio θ_1/θ_2 , and remarkably, it is insensitive to the self-replication and stage-progression rates of the intermediate cells and also the degradation rate of the terminal cell. Observe that this network is an instance of the autocatalytic network consider in Section 4.1.2.

5.3 Phosphorylation cycle based integral controller

Our next example is of a synthetic controller that enables an arbitrary intracellular network to achieve maxRPA by implementing the integral feedback mechanism via a phosphorylation cycle [32] (see Figure 8). This cycle constitutes the IM for the maxRPA network with output $\mathbf{X}_1$, while the network being controlled is the RON. The two species in the cycle are the active phosphorylated form of a substrate $\mathbf{X}_2$ and its inactive unphosphorylated form $\mathbf{X}_2^*$. The rates of switching between these two forms obeys Michaelis-Menten enzyme kinetics, as shown in Table 1. Assuming that both forms of the substrate are in high abundance, the reaction $\mathbf{X}_2^* \rightarrow \mathbf{X}_2$ can be viewed as zeroth-order production while the other reaction $\mathbf{X}_2 \rightarrow \mathbf{X}_2^*$ can be viewed as zeroth-order degradation (see Table 1). Under these approximations, this network becomes a zeroth-order degradation network presented in Section 4.1.1. The IM consists of a single species $\mathbf{X}_2$ and as per Theorem 3.2, the network exhibits maxRPA for the output species $\mathbf{X}_1$ which represents the phosphatase enzyme that is responsible for catalysing the dephosphorylation reaction.

Figure 8: **Integral controller based on phosphorylation cycle.** This figure depicts the scenario, adapted from [32], where an arbitrary network (RON) is controlled by a module (IM) comprising a substrate in phosphorylated X_2^* and unphosphorylated X_2 forms. The rate constant θ_1 for the phosphorylation reaction is determined by the kinase concentration, while the dephosphorylation reaction occurs at a rate that is proportional to the concentration of the phosphatase X_1 which is the RPA output. The arbitrary network (RON) gets actuated by the phosphorylated substrate X_2^* and it produces the phosphatase X_1 .

5.4 The sniffer system

The *sniffer system* [33] (see Figure 9) exemplifies an important regulatory motif called the *incoherent feedforward loop* (IFFL). This motif is found in many biological networks [45] and particularly in signalling networks [48, 49]. In an IFFL the input node influences the output node via two paths that have the opposite effect, i.e. one path activates the production of the output while the other represses it. The structure of IFFL ensures that the cumulative influence of the input node on the output node *cancels out*, thereby ensuring RPA w.r.t all disturbances that only affect the input node.

In the sniffer system shown in Figure 9, the input node is given by species X_3 , species X_1 is the output node, while another species X_2 serves as an intermediate node. The input species X_3 catalytically produces both X_1 and X_2 , and X_2 catalytically degrades X_1 . We include inflow and outflow reactions for X_3 and also an outflow reaction for X_2 . Overall this three-species network has the following six reactions:

The stoichiometry matrix for this network is

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

No.	Name	Reaction	Propensity	Approx. Reaction	Approx. Propensity
1	Phosphorylation	$\mathbf{X}_2^* \longrightarrow \mathbf{X}_2$	$\frac{\theta_1 x_2^*}{x_2^* + K_1}$	$\emptyset \longrightarrow \mathbf{X}_2$	θ_1
2	Dephosphorylation	$\mathbf{X}_2 \longrightarrow \mathbf{X}_2^*$	$\frac{\theta_2 x_1 x_2}{x_2 + K_2}$	$\mathbf{X}_2 \longrightarrow \emptyset$	$\theta_2 x_1$

Table 1: **Reactions in the Phosphorylation cycle.** Both reactions follow Michaelis–Menten kinetics with Michaelis constants K_1 and K_2 , and x_1 , x_2 and x_2^* denote the concentrations of $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_2^*$ respectively. For the phosphorylation reaction, the kinase concentration is fixed and θ_1 is its product with the catalytic rate constant of this enzyme. For the dephosphorylation reaction, θ_1 is the catalytic rate constant of the phosphatase. When x_2 and x_2^* have high values the two reactions and their propensities can be approximated as indicated in the last two columns.

and it can be seen that for $q = (1, 0, 0)$ and $\kappa = 1$, condition (3.17) holds. However the other stoichiometric condition stipulated by Theorem 3.2 does not hold and hence this network will not exhibit maxRPA. This can also be verified independently by noting that the steady-state values of the species concentrations are given by

$$\bar{x}_1 = \frac{\theta_1 \theta_4}{\theta_2 \theta_3}, \quad \bar{x}_2 = \frac{\theta_3 \theta_5}{\theta_4 \theta_6} \quad \text{and} \quad \bar{x}_3 = \frac{\theta_5}{\theta_6}.$$

Hence the output level $\bar{x}_1$ is only robust to perturbations in parameters θ_5 and θ_6 that directly affect the input, while it is sensitive to all the other parameters that determine the flow of information from the input to the output.

6 Conclusion

Intracellular biomolecular networks are immensely complex and yet they must tightly regulate key molecular players — failing to do so can have catastrophic ramifications for the cell-population and consequently for the whole organism [3]. As these networks operate in an unpredictable and fluctuating environment, it is conceivable that they have evolved to attain a form of “maximal robustness”, whereby the output being regulated is sensitive to the least number of networks parameters, and insensitive to maximum number of disturbance sources which can be both parametric or structural. The goal of this paper is to develop a mathematical theory for networks exhibiting *maximal robust perfect adaptation* (maxRPA). We consider both deterministic and stochastic descriptions of the dynamics and provide necessary and sufficient conditions for maxRPA. These conditions pertain to only the stoichiometric structure of the network and they are linear-algebraic in nature, and hence they can be easily checked for large dimensional networks.

Our results are best understood from the control-theoretic viewpoint offered by the well-known *internal model principle* (IMP) (see [28]). It stipulates that for a network to achieve robustness, a subnetwork, called the *internal model* (IM), must measure the deviation of the output from its set-point, and then pass its time-integral as an

Figure 9: **Sniffer system.** This figure depicts the sniffer system, adapted from [33], which comprises an *incoherent feedforward loop* (IFFL) from the input $\mathbf{X}_3$ to the output $\mathbf{X}_1$. The input directly produces the output, but it also produces another species $\mathbf{X}_2$ that represses the output by causing its degradation. The reactions of the network are given by (5.40). All the reactions follow mass-action kinetics with rate constants as indicated in the figure.

actuating signal to the *rest of the network* (RON) in a feedback fashion (see Figure 1). We formalise this principle in the context of maxRPA networks and demonstrate how the IM arises naturally from our stoichiometric conditions that characterise such networks. One can interpret the IM as an embedded controller module that ensures robustness, thereby connecting our results with the synthetic biology paradigm for designing controller modules for generating RPA [8, 24, 32]. An important question is to determine the minimal complexity of the IM that is needed for maxRPA. Using our characterisation results we answer this question fully in both deterministic and stochastic settings and we highlight the differences between the two scenarios. We also illustrate our results with known examples of maxRPA networks in the existing literature.

Existing RPA studies typically assume that disturbances can enter the dynamics via a dedicated input node [18, 19], and in such a setting only two possible RPA network topologies can arise - namely the *incoherent feedforward* (IFF) topology and the *negative feedback* (NFB) topology. In contrast, in this study of maxRPA networks, disturbances can enter from multiple nodes and they can vary the parameters and also alter the network structure. Therefore the set of allowed disturbances is significantly larger, and consequently, maxRPA networks cannot have the IFF topology as here robustness is engendered via a balancing mechanism between disparate output affecting network branches — and this balancing mechanism can be disrupted by the disturbances leading to the loss of RPA (see the example in Section 5.4). As our results show, all maxRPA networks must have the NFB topology, and in addition, the stoi-

chiometric structure of the network must satisfy a certain linear-algebraic condition.

We believe that the notion of maxRPA is important for understanding and designing robust biomolecular networks, especially if robustness needs to be preserved within the cellular milieu which is rife with disturbances. These disturbances include fluctuations in extrinsic factors, such as nutrient concentration, temperature, pressure etc. [1], or intrinsic factors, such as, metabolic burden [21, 22], cross-talk with other cellular components [20], loading effects [23, 40]. In synthetic biology applications, our results can facilitate modular design by providing a framework for building orthogonal components that behave predictably upon connection due to robustness against the connection-induced disturbances [50]. The theory we have developed is in the idealised scenario where the adaptation is perfect and all disturbances that do not affect the output set-point are rejected. However, our results provide a starting point for finding or constructing realistic biological networks that are *nearly* maxRPA. As illustrated by the example in Section 5.3, such networks would *approximately* satisfy the required stoichiometric conditions, under some reasonable assumptions on the abundance levels of the constituent species.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme grant agreement no. 743269 (CyberGenetics project).

References

- [1] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004.
- [2] Mustafa H Khammash. Perfect adaptation in biology. *Cell Systems*, 12(6):509–521, 2021.
- [3] George Morgan Hughes. *Homeostasis and Feedback Mechanisms*. Cambridge University Press, 1964.
- [4] Naama Barkai and Stan Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913, 1997.
- [5] Andre Levchenko and Pablo A Iglesias. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophysical journal*, 82(1):50–63, 2002.
- [6] Dale Muzzey, Carlos A Gómez-Urbe, Jerome T Mettetal, and Alexander van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–171, 2009.
- [7] Tina Toni, Goran Jovanovic, Maxime Huvet, Martin Buck, and Michael PH Stumpf. From qualitative data to quantitative models: analysis of the phage shock protein stress response in escherichia coli. *BMC systems biology*, 5(1):1–16, 2011.

- [8] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. *Cell systems*, 2(1):15–26, 2016.
- [9] Francis Mairet. A biomolecular proportional integral controller based on feedback regulations of protein level and activity. *Royal Society open science*, 5(2):171966, 2018.
- [10] Xiao Yu Ni, Tormod Drengstig, and Peter Ruoff. The control of the controller: molecular mechanisms for robust perfect adaptation and temperature compensation. *Biophysical journal*, 97(5):1244–1253, 2009.
- [11] H El-Samad, JP Goff, and M Khammash. Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective. *Journal of Theoretical Biology*, 214(1):17–29, 2002.
- [12] Daniel M Tveit, Gunhild Fjeld, Tormod Drengstig, Fabian V Filipp, Peter Ruoff, and Kristian Thorsen. Exploring mechanisms of glucose uptake regulation and dilution resistance in growing cancer cells. *bioRxiv*, 2020.
- [13] Avigdor Eldar, Ruslan Dorfman, Daniel Weiss, Hilary Ashe, Ben-Zion Shilo, and Naama Barkai. Robustness of the bmp morphogen gradient in drosophila embryonic patterning. *Nature*, 419(6904):304–308, 2002.
- [14] D.A. Anderson and T.G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In H. Koepl, G. Setti, M. di Bernardo, and D. Densmore, editors, *Design and Analysis of Biomolecular Circuits*. Springer-Verlag, 2011.
- [15] Harley H. McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci., Biochemistry*, 94:814–819, 1997.
- [16] Wenzhe Ma, Ala Trusina, Hana El-Samad, Wendell A Lim, and Chao Tang. Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4):760–773, 2009.
- [17] Zhe F Tang and David R McMillen. Design principles for the analysis and construction of robustly homeostatic biological networks. *Journal of theoretical biology*, 408:274–289, 2016.
- [18] Robyn P Araujo and Lance A Liotta. The topological requirements for robust perfect adaptation in networks of any size. *Nature communications*, 9(1):1757, 2018.
- [19] Yangyang Wang, Zhengyuan Huang, Fernando Antoneli, and Martin Golubitsky. The structure of infinitesimal homeostasis in input–output networks. *Journal of Mathematical Biology*, 82(7):1–43, 2021.
- [20] Stefano Cardinale and Adam Paul Arkin. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. *Biotechnology journal*, 7(7):856–866, 2012.
- [21] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2009.
- [22] Yili Qian, Hsin-Ho Huang, José I Jiménez, and Domitilla Del Vecchio. Resource competition shapes the response of genetic circuits. *ACS synthetic biology*, 6(7):1263–1272, 2017.

- [23] Alejandra C Ventura, Peng Jiang, Lauren Van Wassenhove, Domitilla Del Vecchio, Sofia D Merajver, and Alexander J Ninfa. Signaling properties of a covalent modification cycle are altered by a downstream target. *Proceedings of the National Academy of Sciences*, 107(22):10032–10037, 2010.
- [24] Stephanie K Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, 2019.
- [25] Oren Shoval, Lea Goentoro, Yuval Hart, Avi Mayo, Eduardo Sontag, and Uri Alon. Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences*, 107(36):15995–16000, 2010.
- [26] Corentin Briat, Christoph Zechner, and Mustafa Khammash. Design of a synthetic integral feedback circuit: dynamic analysis and dna implementation. *ACS synthetic biology*, 5(10):1108–1116, 2016.
- [27] Bruce A Francis and William M Wonham. The internal model principle for linear multivariable regulators. *Applied mathematics and optimization*, 2(2):170–194, 1975.
- [28] Bruce A Francis and W Murray Wonham. The internal model principle of control theory. *Automatica*, 12(5):457–465, 1976.
- [29] Eduardo D Sontag. Adaptation and regulation with signal detection implies internal model. *Systems & control letters*, 50(2):119–126, 2003.
- [30] Tau-Mu Yi, Yun Huang, Melvin I Simon, and John Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences*, 97(9):4649–4653, 2000.
- [31] Arthur D Lander, Kimberly K Gokoffski, Frederic Y M Wan, Qing Nie, and Anne L Calof. Cell lineages and the logic of proliferative control. *PLoS biology*, 7(1):e1000015, 2009.
- [32] Hsin-Ho Huang, Yili Qian, and Domitilla Del Vecchio. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nature Communications*, 9(1):5415, 2018.
- [33] John J Tyson, Katherine C Chen, and Bela Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current opinion in cell biology*, 15(2):221–231, 2003.
- [34] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.
- [35] Ankit Gupta and Mustafa Khammash. Computational identification of irreducible state-spaces for stochastic reaction networks. *SIAM Journal on Applied Dynamical Systems*, 17(2):1213–1266, 2018.
- [36] Ankit Gupta, Corentin Briat, and Mustafa Khammash. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS Comput Biol*, 10(6):e1003669, 06 2014.
- [37] Corentin Briat. A biology-inspired approach to the positive integral control of positive systems: The antithetic, exponential, and logistic integral controllers. *SIAM Journal on Applied Dynamical Systems*, 19(1):619–664, 2020.

- [38] Sean Meyn. *Control techniques for complex networks*. Cambridge University Press, Cambridge, 2008.
- [39] Peter W Glynn and Sean P Meyn. A liapounov bound for solutions of the poisson equation. *The Annals of Probability*, pages 916–931, 1996.
- [40] Domitilla Del Vecchio. Modularity, context-dependence, and insulation in engineered biological circuits. *Trends in biotechnology*, 33(2):111–119, 2015.
- [41] Guy Shinar and Martin Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.
- [42] Daniele Cappelletti, Ankit Gupta, and Mustafa Khammash. A hidden integral structure endows absolute concentration robust systems with resilience to dynamical concentration disturbances. *Journal of the Royal Society Interface*, 17(171):20200437, 2020.
- [43] Deepak K Agrawal, Ryan Marshall, Vincent Noireaux, and Eduardo D Sontag. In vitro implementation of robust gene regulation in a synthetic biomolecular integral controller. *Nature communications*, 10(1):1–12, 2019.
- [44] T Frei, C-H Chang, M Filo, and M Khammash. Genetically engineered integral feedback controllers for robust perfect adaptation in mammalian cells. *bioRxiv*, 2020.
- [45] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.
- [46] Claudio D Stern and Scott E Fraser. Tracing the lineage of tracing cell lineages. *Nature cell biology*, 3(9):E216–E218, 2001.
- [47] Markus Loeffler and Ingo Roeder. Tissue stem cells: definition, plasticity, heterogeneity, self-organization and models—a conceptual approach. *Cells Tissues Organs*, 171(1):8–26, 2002.
- [48] James E Ferrell. Perfect and near-perfect adaptation in cell signaling. *Cell systems*, 2(2):62–67, 2016.
- [49] Satoru Sasagawa, Yu-ichi Ozaki, Kazuhiro Fujita, and Shinya Kuroda. Prediction and validation of the distinct dynamics of transient and sustained erk activation. *Nature cell biology*, 7(4):365–373, 2005.
- [50] Theodore W Grunberg and Domitilla Del Vecchio. Modular analysis and design of biological circuits. *Current opinion in biotechnology*, 63:41–47, 2020.