# Context in synthetic biology: Memory effects of environments with mono-molecular reactions

View Online      Export Citation      CrossMark

Johannes Falk,[1,a] (iD)   Leo Bronstein,[2,3]   Maleen Hanst,[2]   Barbara Drossel,[1] and Heinz Koeppl[2,b]

AFFILIATIONS

[1] Institut für Festkörperphysik, Technische Universität Darmstadt, Hochschulstr. 6, 64289 Darmstadt, Germany
[2] Bioinspired Communication Systems, Technische Universität Darmstadt, Rundeturmstr. 12, 64283 Darmstadt, Germany
[3] Institute of Biostatistics and Clinical Research, University of Münster, Schmeddingstr. 56, 48149 Münster, Germany

[a] Electronic mail: falk@fkp.tu-darmstadt.de
[b] Electronic mail: heinz.koeppl@bcs.tu-darmstadt.de

## ABSTRACT

Synthetic biology aims at designing modular genetic circuits that can be assembled according to the desired function. When embedded in a cell, a circuit module becomes a small subnetwork within a larger environmental network, and its dynamics is therefore affected by potentially unknown interactions with the environment. It is well-known that the presence of the environment not only causes extrinsic noise but also memory effects, which means that the dynamics of the subnetwork is affected by its past states via a memory function that is characteristic of the environment. We study several generic scenarios for the coupling between a small module and a larger environment, with the environment consisting of a chain of mono-molecular reactions. By mapping the dynamics of this coupled system onto random walks, we are able to give exact analytical expressions for the arising memory functions. Hence, our results give insights into the possible types of memory functions and thereby help to better predict subnetwork dynamics.

*Published under license by AIP Publishing.* https://doi.org/10.1063/1.5053816

## I. INTRODUCTION

One of the key goals of synthetic biology is the development of modular genetic devices that can be assembled in a flexible manner according to the desired task.[1] Each device is designed to provide exactly one function in a robust and reliable fashion, analogous to, e.g., resistors and capacitors in electrical circuits. Conventionally, development and optimization of each module are done in isolation, and the output parameters are analyzed under controlled conditions. Based on this characterization, complex genetic circuits of several interconnected modules are designed, optimized, and implemented. However, upon interconnection and implementation into a cell, each circuit finds itself embedded in a fluctuating environment that possibly affects the intended functioning.[2] In the field of synthetic biology, these environmental influences are often subsumed as context effects and can, according to their origin, be subdivided into different categories:[3,4] *compositional* context describes the perturbation of a module

due to the functional composition of the device, *host* context captures the influences provided by the cellular environment, and *external* context captures all disturbances that originate outside the cell. In order to design reliable modules, it is essential to analyze and subsequently robustify network designs specifically against these different and often unavoidable contextual effects. Consequently, several authors have studied the impact of context effects and proposed methods to account for them already during the design process of genetic devices.

Saez-Rodriguez *et al.* and later Del Vecchio *et al.* considered compositional context effects by investigating how the addition of a downstream module can change the output of the upstream module. This effect was coined *retroactivity*.[5,6] The authors proposed a general approach to quantifying the effects of retroactivity. Additionally, they designed an insulating device that attenuates retroactivity and later successfully demonstrated its functionality. A basic assumption for

Del Vecchio's analysis is that there is a separation of time scales between the dynamics of the two described modules. While this assumption is often valid for compositional context effects, it cannot be applied to host context effects since the cellular environment often has similar time scales to the designed subnetwork. In spite of this, the interactions with the host environment are often ignored by assuming that the copy numbers of environmental species are very large or the environment is efficiently buffered. An important step forward was made by Liebermeister *et al.* who developed a general framework that can be used to include compositional as well as host context effects into the rate equations, especially accounting for the dynamical response of the environment to the output of the subnetwork.[7] To this purpose, they divided the full network into a network of interest, the so-called subnetwork, and the surrounding environment. After linearizing the equations for the environmental variables around a steady state, they reduce its dimensionality by projecting the dynamics on the subspace of the dominant dynamic modes. Applying their method to a metabolic network embedded in an environment with parameter uncertainty, Liebermeister *et al.* demonstrated how accounting for the environment improves the predictions of the model, even if the parameters are not known exactly. While model reduction via balanced truncation was shown to outperform conventional approaches based on fixed environmental concentrations, Liebermeister *et al.* also mentioned several limitations of their method, namely, the requirement for a steady state in the environment and the non-preservation of conservation relations between the subnet and the environment. The results of Liebermeister *et al.* were recently extended by Rubin *et al.* who used the Mori-Zwanzig projection operator formalism to obtain a description for a generic subnetwork that includes the contextual effects via the so-called memory terms.[8] By using memory terms, it is possible to include complex mutual interaction between the subnetwork and the environment into the description of the subnetwork, for instance, when the subnetwork influences the environment and these environmental changes are coupled back to the subnetwork later in time. Using the projection technique, the authors demonstrated that the inclusion of memory terms leads to higher accuracy in the predicted subnetwork dynamics.

A different approach to including environmental effects into the description of the subnet was pursued by Zechner *et al.*[9,10] and Bronstein and Koeppl.[11] In the description of the stochastic process for the evolution of the joint network, the environmental states were integrated out, resulting in a description of the subnet that uses the best estimate of the environmental state, based on the past observations of the subnet.

In this paper, we will extend these various investigations of environmental effects on the subnet in several respects. In contrast to Rubin *et al.*, who based their derivation on the chemical Fokker-Planck equation (CFPE) in the limit of vanishing noise, we will apply the Nakajima-Mori-Zwanzig projection operator framework directly to the chemical master equation (CME) to marginalize out the environmental states

since the CME is the more fundamental equation for stochastic modeling of a chemical reaction network. We will show that for a reaction network with linear environment and linear subnet-environment interactions, the marginal description of the CME still possesses closed first-order moment equations, and that these coincide with the equations obtained by applying the projection framework to an ordinary differential equation (ODE) model of the same form as the reaction rate equation (RRE). While gene-regulatory networks are usually nonlinear, in certain situations, they can be simplified to linear networks either by expansion around a steady-state[7] or by mapping them on approximately equivalent systems.[12]

Based on this linear description, a main goal of our paper is the systematic study of the influence of generic environments on subnetworks. In distinction to previous studies of marginalization,[8] we focus on the situation that we do not have specific knowledge of the environment into which a module might be embedded. By considering several chain-like prototypical environments (e.g., cascaded genetic module environments), we derive analytical expressions for the memory kernels of these common cellular environmental structures. Those environmental building blocks can be used as general-purpose environment models when investigating the sensitivity of a module to contextual influences. To demonstrate the effects of such contextual influences, we take, similarly to Del Vecchio *et al.*, the example of an oscillating genetic module that interacts with an environment of mono-molecular reactions. Additionally, we demonstrate a more complex context effect using the example of a negative-feedback module that drives interconnected downstream modules. This shows that even environments that seem to be too simple to be relevant considerably change the dynamics of a module. Furthermore, our analytical results can be extended to include a broad class of environments into the analysis of synthetic genetic networks.

The outline of the remainder of the paper is as follows: In Sec. II, we derive stochastic and deterministic descriptions for the subnet that contain the effects of the environment via memory functions and relate the corresponding marginal description. Section III presents derivations of exact analytical expressions for the memory terms of prototypical environments, which will be based on the deterministic marginal description. Finally in Sec. IV, we use the memory functions of several simple model environments and demonstrate their effect on two specific subnets.

## II. MODEL AND METHODS

### A. Reaction network: System and environment

In the following, we call the system of interest the *subnetwork* or *target network*, indicating that it is part of a larger network that comprises the system of interest and its environment. We model the full system as a reaction network consisting of $N + M$ species, where $N$ and $M$ denote the number of subnet and environmental species, respectively, and $R$

reactions,

$$\sum_{n=1}^{N} s_{nj} X_n^S + \sum_{n=1}^{M} s'_{nj} X_n^E \longrightarrow \sum_{n=1}^{N} r_{nj} X_n^S + \sum_{n=1}^{M} r'_{nj} X_n^E$$
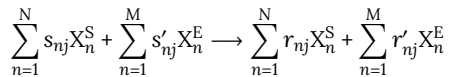
for $j = 1, \ldots, R$. The superscripts S and E denote the subnet and environmental species, respectively. The coefficients $s_{1j}, \ldots, s_{Nj}$ and $s'_{1j}, \ldots, s'_{Mj}$ are the stoichiometric substrate coefficients, and $r_{1j}, \ldots, r_{Nj}$ and $r'_{1j}, \ldots, r'_{Mj}$ are the stoichiometric product coefficients of reaction $j$.

## B. Stochastic and deterministic dynamics

For a well-mixed system, the time evolution of the network is modeled by a continuous-time Markov chain $\mathbf{X}(t) = (\mathbf{X}^S(t), \mathbf{X}^E(t))$ in which the state $\boldsymbol{x} = (\boldsymbol{x}^S, \boldsymbol{x}^E) \in \mathbb{N}_0^N \times \mathbb{N}_0^M$ of the system specifies the (integer-valued) copy number of each species. The probability $p(t, \boldsymbol{x}) := P(\mathbf{X}(t) = \boldsymbol{x})$ evolves according to the CME

$$\frac{\mathrm{d}}{\mathrm{d}t} p(t, \boldsymbol{x}) = \mathsf{T} p(t, \boldsymbol{x}) \tag{1}$$

with forward evolution operator

$$[\mathsf{T} p](\boldsymbol{x}) = \sum_{j=1}^{R} \left\{ h_j(\boldsymbol{x} - \boldsymbol{\nu}_j) p(\boldsymbol{x} - \boldsymbol{\nu}_j) - h_j(\boldsymbol{x}) p(\boldsymbol{x}) \right\}. \tag{2}$$

Here, $h_j(\boldsymbol{x})$ is the rate function of reaction $j$ when the system is in state $\boldsymbol{x}$, also called hazard function, and $\boldsymbol{\nu}_j = (\boldsymbol{\nu}_j^S, \boldsymbol{\nu}_j^E)$ is the stoichiometric change vector of reaction $j$ with $\boldsymbol{\nu}_j^S = (r_{1j} - s_{1j}, \ldots, r_{Nj} - s_{Nj})$ and $\boldsymbol{\nu}_j^E = (r'_{1j} - s'_{1j}, \ldots, r'_{Mj} - s'_{Mj})$.

We say that a stochastic reaction network has *mass-action kinetics* if its hazards are given by

$$h_j(\boldsymbol{x}) = \tilde{c}_j \prod_{i=1}^{N} \binom{\boldsymbol{x}_i^S}{s_{ij}} \prod_{l=1}^{M} \binom{\boldsymbol{x}_l^E}{s'_{lj}}$$

for $j = 1, \ldots, R$, where $\tilde{c}_j$ is a constant rate parameter.

A reaction is said to be linear if its hazard function is linear. For a system with mass-action kinetics, a reaction is linear if at most one stoichiometric substrate coefficient is nonzero, i.e., a linear reaction takes one of the two forms

$$\emptyset \longrightarrow [\cdots], \qquad X_i \longrightarrow [\cdots],$$

where the right-hand side of the reactions can be arbitrary. A reaction is said to be mono-molecular if at most one stoichiometric substrate as well as one stoichiometric product coefficient is nonzero, i.e., it takes one of the forms

$$\emptyset \longrightarrow X_i, \quad X_i \longrightarrow \emptyset, \quad X_i \longrightarrow X_j,$$
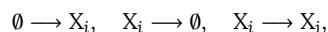
where the first reaction type is called a birth, the second called a degradation, and the third called a conversion reaction. Every mono-molecular reaction is linear.

Finally, a reaction network is termed linear (respectively, mono-molecular) if all reactions are linear (respectively, mono-molecular). In particular, we define the environment

as linear (respectively, mono-molecular) if all environment-environment interactions are linear (respectively, mono-molecular).

Furthermore, recall that for a function $\psi(\boldsymbol{x})$ the evolution equation for the moment $\langle \psi \rangle_t = \sum_{\boldsymbol{x}} p(t, \boldsymbol{x}) \psi(\boldsymbol{x})$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \psi \rangle_t = \sum_{\boldsymbol{x}} \psi(\boldsymbol{x}) \mathsf{T} p(t, \boldsymbol{x}) = \sum_{\boldsymbol{x}} [\mathsf{T}^\dagger \psi](\boldsymbol{x}) p(t, \boldsymbol{x})$$
$$= \langle \mathsf{T}^\dagger \psi \rangle_t, \tag{3}$$

where $\mathsf{T}^\dagger$, given by

$$[\mathsf{T}^\dagger \psi](\boldsymbol{x}) = \sum_{j=1}^{R} h_j(\boldsymbol{x}) \left\{ \psi(\boldsymbol{x} + \boldsymbol{\nu}_j) - \psi(\boldsymbol{x}) \right\},$$

denotes the adjoint of the operator $\mathsf{T}$ with respect to the bilinear form $(p, \psi) := \sum_{\boldsymbol{x}} p(\boldsymbol{x}) \psi(\boldsymbol{x})$. For a linear reaction network with mass-action kinetics, the corresponding first-order moment equations (the *mean equations*) are closed and linear

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \boldsymbol{x} \rangle_t = \sum_{j=1}^{R} \langle h_j(\boldsymbol{x}_t) \rangle \boldsymbol{\nu}_j = \sum_{j=1}^{R} h_j(\langle \boldsymbol{x} \rangle_t) \boldsymbol{\nu}_j = \mathbf{A} \langle \boldsymbol{x} \rangle_t + \mathbf{b},$$

where the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are defined by this equation. If the network is non-linear, the mean equations are not closed.

At this point, we turn to considering deterministic dynamics of reaction networks. One possibility to obtain a deterministic description of the dynamics is to start with the continuous-time Markov chain model and take the large volume limit where the reaction rate equations are recovered. To achieve this, we introduce the concentration variable $\boldsymbol{\mu} := \mathbf{X}(t)/V$, where $V$ is the system size. Inserting this concentration variable in the corresponding process representation of a continuous-time Markov chain, namely, in the random time-change representation of Kurtz[13] and applying the law of large numbers, one obtains the well-known reaction rate equation (RRE)

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{\mu}(t) = \sum_{j=1}^{R} \lambda_j(\boldsymbol{\mu}(t)) \, \boldsymbol{\nu}_j. \tag{4}$$

Here, the rate functions $\lambda_j$ are called mass-action propensities for concentrations, and in terms of concentrations for subnet and environmental species, $\boldsymbol{\mu} = (\boldsymbol{\mu}^S, \boldsymbol{\mu}^E)$, they are given by

$$\lambda_j(\boldsymbol{\mu}^S, \boldsymbol{\mu}^E) = c_j \prod_{i=1}^{N} (\mu_i^S)^{s_{ij}} \prod_{l=1}^{M} (\mu_l^E)^{s'_{lj}}$$

with the macroscopic rate constants $c_j$. Hence, for a reaction system with high copy numbers or high concentrations, the dynamics can be well described by the deterministic dynamics, the RRE.

## C. The projection operator formalism

We use the projection operator formalism for both the stochastic and deterministic description to derive effective

equations for the dynamics of the subnet. For the stochastic description, we apply the formalism[14] (compare also the original work of Nakajima,[15] Zwanzig,[16] and Mori[17]) to derive a generalized master equation. This equation describes the evolution of the marginal probability distribution.

For the CME, a natural definition of a marginalization operator is

$$[Mp](\mathbf{x}^S) = \sum_{\mathbf{x}^E} p(\mathbf{x}^S, \mathbf{x}^E). \tag{5}$$

This operator can be used within the projection operator framework as described in the following.

At first, we observe that the marginalization operator maps the distribution of the joint state space of the subnet and environment to the marginal distribution over the subnet state space, whereas the projection operator framework requires a projection operator that maps the space of distributions over the joint space to itself. Therefore, in addition to the marginalization operator M defined in (5), a lifting operator L acting on distributions $p(\mathbf{x}^S)$ over the subnet states has to be defined. We choose[18,19]

$$[Lp](\mathbf{x}^S, \mathbf{x}^E) = q(\mathbf{x}^E)p(\mathbf{x}^S)$$

for some fixed distribution $q$ over the environmental states.

The (time-independent) operator $P := LM$ is a projection, i.e., we have $P^2 = P$. Defining the orthogonal projection operator $Q := I - P$, we also have

$$Q^2 = Q, \quad PQ = QP = 0.$$

Applying M (respectively, Q) to the CME (1) and using $P + Q = I$ and $P = LM$, we obtain the following two equations:

$$\frac{d}{dt}Mp = MTLMp + MTQp, \tag{6}$$

$$\frac{d}{dt}Qp = QTQp + QTLMp. \tag{7}$$

Note that the projected distribution $p^S = Mp$ fully contains the exact subnet dynamics due to the special form of our projection operator.

Formally solving (7) results in

$$Qp(t) = e^{tQT}Qp(0) + \int_0^t dt' e^{(t-t')QT}QTLp^S(t'),$$

and inserting the latter into (6) finally yields

$$\frac{d}{dt}p^S(t) = MTLp^S(t) + \int_0^t dt' MTe^{(t-t')QT}QTLp^S(t') + MTe^{tQT}Qp(0). \tag{8}$$

Here, the first term reflects the Markovian part of the dynamics and the second (convolution) term the *memory*. The third term represents a *noise* since we assume to have no knowledge about the initial distribution of the environment. The noise

term can be made to vanish if we assume that Q is orthogonal to the initial distribution.

Equation (8) has a reduced number of dimensions compared to the full network, but the description has not become simpler since the equation still constitutes the exact dynamics of the subnet including the effects of the environment on it—reflected by the memory and noise terms. To obtain simpler expressions, one needs in general approximations, especially for the memory kernel. However, in this paper we restrict our considerations to networks with mono-molecular within-environment as well as subnet-environment interactions, for which an analytical treatment of prototypical environments is possible. We can then decompose the full time-evolution operator T as $T = T_0 + T_1$, where $T_0$ includes all linear reactions and $T_1$ contains the non-linear subnet-subnet interactions. Strictly speaking, the operator $T_1$ acts on distributions over the joint space of the subnet and environment. However, in the following, we will sometimes implicitly consider it to act on distributions over the subnet state only.

### D. Mean equation for the projected CME

In this section, we extract marginal mean equations from the marginal CME (8). Starting from (8), the expectation $\langle \phi \rangle$ of any function $\phi(\mathbf{x}^S)$ evolves according to

$$\frac{d}{dt}\langle \phi \rangle_t = \left\langle L^\dagger T^\dagger M^\dagger \phi \right\rangle_t + \int_0^t dt' \left\langle L^\dagger T^\dagger Q^\dagger e^{(t-t')T^\dagger Q^\dagger} T^\dagger M^\dagger \phi \right\rangle_{t'} + \left\langle Q^\dagger e^{tT^\dagger Q^\dagger} T^\dagger M^\dagger \phi \right\rangle_0. \tag{9}$$

Thus, we need expressions for the adjoints of the operators involved. Let $\psi(\mathbf{x}^S, \mathbf{x}^E)$ and $\phi(\mathbf{x}^S)$ be two arbitrary functions. A brief computation shows that we have

$$[M^\dagger \phi](\mathbf{x}^S, \mathbf{x}^E) = \phi(\mathbf{x}^S),$$
$$[L^\dagger \psi](\mathbf{x}^S) = \sum_{\mathbf{x}^E} q(\mathbf{x}^E)\psi(\mathbf{x}^S, \mathbf{x}^E),$$

and thus

$$[P^\dagger \psi](\mathbf{x}^S, \mathbf{x}^E) = \sum_{\hat{\mathbf{x}}^E} q(\hat{\mathbf{x}}^E)\psi(\mathbf{x}^S, \hat{\mathbf{x}}^E).$$

Choosing $\boldsymbol{\phi}(\mathbf{x}^S) = \mathbf{x}^S$, we obtain from (9) equations for the mean abundances. To get an explicit expression, we first consider the case of a fully linear network. For a linear reaction network, all reactions are of one of the two forms

$$\emptyset \longrightarrow [\cdots], \qquad X_i \longrightarrow [\cdots],$$

where the first form can be seen as a special case of the second form by including an auxiliary species $X_0$ with abundance 1 and replacing the reaction

$$\emptyset \longrightarrow [\cdots] \quad \text{by} \quad X_0 \longrightarrow X_0 + [\cdots].$$
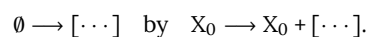
We will assume in the following that this rewriting has been performed.

Noting that $\phi$ is linear, we can obtain these equations in a more explicit form by first verifying that each operator involved in (9) leaves the space of affine-linear functions invariant. Then (9), which is infinite-dimensional, reduces to a finite-dimensional equation for the coefficients of an affine-linear function. Thus, let

$$\psi(\boldsymbol{x}^S, \boldsymbol{x}^E) = \boldsymbol{w}^S \boldsymbol{x}^S + \boldsymbol{w}^E \boldsymbol{x}^E + v, \quad \phi(\boldsymbol{w}^S) = \boldsymbol{w}^S \boldsymbol{x}^S + v$$

be general affine-linear functions, where $\boldsymbol{w}^S$ and $\boldsymbol{w}^E$ are row vectors and $v$ is a scalar. As explained in Sec. II B, the first-order moment equations for a linear network are closed. The action of the operator $\mathsf{T}_0^\dagger$ can then be specified via the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{SS} & \mathbf{A}^{SE} \\ \mathbf{A}^{ES} & \mathbf{A}^{EE} \end{bmatrix}$$

as

$$[\mathsf{T}_0^\dagger \psi](\boldsymbol{x}^S, \boldsymbol{x}^E) = [\boldsymbol{w}^S, \boldsymbol{w}^E] \mathbf{A} \begin{bmatrix} \boldsymbol{x}^S \\ \boldsymbol{x}^E \end{bmatrix}.$$

Similarly,

$$[\mathsf{M}^\dagger \phi](\boldsymbol{x}^S, \boldsymbol{x}^E) = \boldsymbol{w}^S \boldsymbol{x}^S + v,$$
$$[\mathsf{L}^\dagger \psi](\boldsymbol{x}^S) = \boldsymbol{w}^S \boldsymbol{x}^S + \boldsymbol{w}^E \langle \boldsymbol{x}^E \rangle_q + v,$$
$$[\mathsf{P}^\dagger \psi](\boldsymbol{x}^S, \boldsymbol{x}^E) = \boldsymbol{w}^S \boldsymbol{x}^S + \boldsymbol{w}^E \langle \boldsymbol{x}^E \rangle_q + v,$$
$$[\mathsf{Q}^\dagger \psi](\boldsymbol{x}^S, \boldsymbol{x}^E) = \boldsymbol{w}^E (\boldsymbol{x}^E - \langle \boldsymbol{x}^E \rangle_q),$$
$$[\mathsf{T}_0^\dagger \mathsf{Q}^\dagger \psi](\boldsymbol{x}^S, \boldsymbol{x}^E) = (\boldsymbol{w}^E \mathbf{A}^{ES}) \boldsymbol{x}^S + (\boldsymbol{w}^E \mathbf{A}^{EE}) \boldsymbol{x}^E.$$

Thus, each operator leaves the space of affine-linear functions invariant. Additionally, the operator $\mathsf{T}_0^\dagger$ removes the inhomogeneous term $v$.

We can now evaluate each of the three terms in (9) in turn. For the Markovian part, we obtain

$$\left\langle \mathsf{L}^\dagger \mathsf{T}^\dagger \mathsf{M}^\dagger \boldsymbol{\phi} \right\rangle_t = \left\langle \mathsf{T}_1^\dagger \boldsymbol{\phi} \right\rangle_t + \mathbf{A}^{SS} \boldsymbol{\mu}^S + \mathbf{A}^{SE} \langle \boldsymbol{x}^E \rangle_q. \tag{10}$$

To evaluate the memory term, we first note that, because $\mathsf{M}^\dagger \phi$ does not depend on $\boldsymbol{x}^E$, we have $\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{M}^\dagger \phi = \mathsf{Q}^\dagger \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi$. The latter does not depend on $\boldsymbol{x}^S$ so that we obtain $\mathsf{T}^\dagger \mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{M}^\dagger \phi = \mathsf{T}_0^\dagger \mathsf{Q}^\dagger \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi$. Similarly, one can verify that $\mathsf{T}^\dagger \mathsf{Q}^\dagger \phi = \mathsf{T}_0^\dagger \mathsf{Q}^\dagger \phi$. For the memory term of (9), we note that the coefficients of the linear function $e^{(t-t')\mathsf{T}^\dagger \mathsf{Q}^\dagger} \mathsf{T}^\dagger \mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{M}^\dagger \phi = e^{(t-t')\mathsf{T}_0^\dagger \mathsf{Q}^\dagger} \mathsf{T}_0^\dagger \mathsf{Q}^\dagger \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi$ are the solution of the ODE system

$$\dot{\boldsymbol{w}}^S = \boldsymbol{w}^E \mathbf{A}^{ES},$$
$$\dot{\boldsymbol{w}}^E = \boldsymbol{w}^E \mathbf{A}^{EE},$$

at time $t - t'$ with initial conditions at time $0$ given by the coefficients of the linear function $\mathsf{T}_0^\dagger \mathsf{Q}^\dagger \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi$. Combining all

of these, we obtain for the memory term

$$\int_0^t dt' \left\langle \mathsf{L}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger e^{(t-t')\mathsf{T}^\dagger \mathsf{Q}^\dagger} \mathsf{T}^\dagger \mathsf{M}^\dagger \boldsymbol{\phi} \right\rangle_{t'}$$
$$= \int_0^t dt' \mathbf{A}^{SE} e^{(t-t')\mathbf{A}^{EE}} \left\{ \mathbf{A}^{ES} \boldsymbol{\mu}^S(t') + \mathbf{A}^{EE} \langle \boldsymbol{x}^E \rangle_q \right\}$$
$$= \int_0^t dt' \mathbf{A}^{SE} e^{(t-t')\mathbf{A}^{EE}} \mathbf{A}^{ES} \boldsymbol{\mu}^S(t')$$
$$+ \mathbf{A}^{SE} e^{t\mathbf{A}^{EE}} \langle \boldsymbol{x}^E \rangle_q - \mathbf{A}^{SE} \langle \boldsymbol{x}^E \rangle_q.$$

For the noise term in (9), we use the relation $\mathsf{Q}^\dagger e^{t\mathsf{T}^\dagger \mathsf{Q}^\dagger} = \mathsf{Q}^\dagger e^{t\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger}$. Since $\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger \mathsf{T}_1^\dagger \mathsf{M}^\dagger \phi = 0$, we have

$$\mathsf{Q}^\dagger e^{t\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger} \mathsf{T}_1^\dagger \mathsf{M}^\dagger \phi = \mathsf{Q}^\dagger \mathsf{T}_1^\dagger \mathsf{M}^\dagger \phi = 0.$$

We then find

$$\left\langle \mathsf{Q}^\dagger e^{t\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger} \mathsf{T}^\dagger \mathsf{M}^\dagger \phi \right\rangle_0 = \left\langle \mathsf{Q}^\dagger e^{t\mathsf{Q}^\dagger \mathsf{T}^\dagger \mathsf{Q}^\dagger} \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi \right\rangle_0$$
$$= \left\langle \mathsf{Q}^\dagger e^{t\mathsf{Q}^\dagger \mathsf{T}_0^\dagger \mathsf{Q}^\dagger} \mathsf{T}_0^\dagger \mathsf{M}^\dagger \phi \right\rangle_0$$
$$= 0.$$

The full equation for the means thus reads

$$\frac{d}{dt} \langle \boldsymbol{x}^S \rangle_t = \langle \mathsf{T}_1^\dagger \boldsymbol{x}^S \rangle_t + \mathbf{A}^{SS} \langle \boldsymbol{x}^S \rangle_t + \int_0^t dt' \mathbf{A}^{SE} e^{(t-t')\mathbf{A}^{EE}} \mathbf{A}^{ES} \langle \boldsymbol{x}^S \rangle_{t'}$$
$$+ \mathbf{A}^{SE} e^{t\mathbf{A}^{EE}} \langle \boldsymbol{x}^E \rangle_q. \tag{11}$$

We see that the non-linearity of the subnet enters only in the Markovian part.

At this point, we proceed to relate the projection operator formalism for the CME as given by (8) and the corresponding mean equation (11) to the formalism applied to the ODE model, as given by the reaction rate equation (4). In particular, we observe that in the case of linear subnet-environment and intra-environment interactions, the RRE (4) takes the form

$$\frac{d}{dt} \boldsymbol{\mu}^S = f(\boldsymbol{\mu}^S) + \mathbf{A}^{SE} \boldsymbol{\mu}^E,$$
$$\frac{d}{dt} \boldsymbol{\mu}^E = \mathbf{A}^{EE} \boldsymbol{\mu}^E + \mathbf{A}^{ES} \boldsymbol{\mu}^S + \mathbf{b}^E. \tag{12}$$

Here, $f$ specifies the subnet-subnet interactions, which can be arbitrary and, in particular, non-linear. Due to the linearity of the remaining reactions, the corresponding dynamics can be represented via (coupling) matrices $\mathbf{A}$, with entries consisting of rate constants $\{c_j\}$, and the vector $\mathbf{b}^E$, containing the rate constants of the environmental birth reactions.

Similarly as for the CME, we can derive an effective (projected) subnet description for (12). We solve the equation for $\boldsymbol{\mu}^E$ and plug the result into the equation for $\boldsymbol{\mu}^S$. The solution reads

$$\dot{\boldsymbol{\mu}}^S = f(\boldsymbol{\mu}^S) + \int_0^t dt' \mathbf{A}^{SE} e^{(t-t')\mathbf{A}^{EE}} (\mathbf{A}^{ES} \boldsymbol{\mu}^S(t') + \mathbf{b}^E) + \mathbf{A}^{SE} e^{t\mathbf{A}^{EE}} \boldsymbol{\mu}^E(0). \tag{13}$$

This is the marginal mean equation as given by (11) if we choose the distribution $q$ to have the correct mean, $\langle \mathbf{x}^E \rangle_q = \boldsymbol{\mu}^E(0)$. This demonstrates that the marginal CME and the marginal equations obtained from (12) are consistent. Interestingly though, in our derivation above, the Markovian, memory and noise terms in (9) and (11) do not correspond to each other directly. While the noise term in (9) vanishes, the memory term in (9) includes $\mathbf{A}^{SE} e^{t\mathbf{A}^{EE}} \langle \mathbf{x}^E \rangle_q$ which is the noise term in (11) since it depends on the conditions of the environment which are unknown in general. It also corresponds to the noise term in (13) if $q$ is chosen as stated above, $\langle \mathbf{x}^E \rangle_q = \boldsymbol{\mu}^E(0)$. In addition, the noise term in (9) vanishes in the first-order moment equations even when the initial distribution $p_0(\mathbf{x}^S, \mathbf{x}^E)$ does not factorize into a product $p_0(\mathbf{x}^S)p_0(\mathbf{x}^E)$.

Since (13) will be used in Sec. III below for the investigation of memory effects in synthetic biology, we want to give some intuition into the different terms of (13), similarly as for (8). The first term $f(\boldsymbol{\mu}^S)$ describes the dynamics that is generated by all the reactions that have only reactants from inside the subnetwork. We will hence call this term in Secs. III and IV the *local dynamics* part.

The second term $\mathbf{A}^{SE} e^{t\mathbf{A}^{EE}} \boldsymbol{\mu}^E(0) =: \boldsymbol{r}(t)$ is the contribution that originates in the (often unknown) initial condition of the environment. Since this term cannot be affected by the subnetwork itself and depends solely on the initial conditions, this term is coined *noise* term.

The third term

$$\int_0^t \mathrm{d}t' \mathbf{M}(t-t') \boldsymbol{\mu}^S(t'), \quad \mathbf{M}(t-t') = \mathbf{A}^{SE} e^{(t-t')\mathbf{A}^{EE}} \mathbf{A}^{ES} \quad (14)$$

is the *memory term*, where $\mathbf{M}(t-t')$ denotes the so-called *memory kernel* that depends on the dynamics happening inside the environment during the time interval $t-t'$. This term is due to the fact that the past states of the subnetwork affected the time evolution of the environment, which now acts back on the subnetwork. The matrices $\mathbf{A}^{SE}$ and $\mathbf{A}^{ES}$ are the coupling matrices from the environment to the subnetwork and from the subnetwork to the environment, respectively. In contrast to the noise term, the influence of the memory term on the subnetwork dynamics is therefore predictable upon knowledge of the structure of the environment.

Although in this work we focus on the marginal mean equation as a first measure to quantify the dynamical behavior of the subnetwork, one can similarly derive the marginal second moment equation of the subnetwork for a completely linear reaction network (see the Appendix).

## III. PROTOTYPIC ENVIRONMENTS WITH MONO-MOLECULAR REACTIONS

While the full environmental network typically has an unknown and complex structure, the interactions between the subnetwork and the environmental network can be simple, e.g., the binding of a subnetwork species to an environmental promoter and vice versa, or constant inflow and outflow of species. Proteins can also be bound in more complex topologies like moiety-conserved cycles or temporarily evolve into promiscuous conformers.[20-22] Likewise, RNA-folding can lead to transitions between different transient structures, where only one is functional.[23] These parasitic interactions and the possibly resulting memory terms need to be considered when designing and evaluating synthetic genetic networks. Thus, it is desirable to obtain memory kernels for some prototypical environments. To this purpose, we will analytically derive memory functions for the broad class of chain-like reaction topologies with mono-molecular and linear interactions, as depicted in Fig. 1. Since in biology almost every reaction is bidirectional (where one direction is conventionally preferred), we include possible reverse reactions in our environments. Additionally, each species has a constant in- and linear outflow due to birth and death reactions. Such a system is sketched in Fig. 1, where $\sigma$ is the rate of inflow at each node, $\kappa$ is the decay rate, and $\alpha, \beta$ are the rates on the connections along the chain. Before we derive the corresponding memory functions, we will show that a chain only exposed to constant inflow is (after an initial transient time, which we do not consider) in a steady state and is hence captured by the noise term $r(t)$ that does not affect the memory terms. The noise term can be obtained by solving the following system of coupled linear equations:

$$\begin{aligned}
x_1 &= \frac{\sigma + x_2 \beta}{\alpha + \kappa}, \\
x_m &= \frac{\sigma + \alpha x_{m-1} + \beta x_{m+1}}{\alpha + \beta + \kappa}, \\
x_M &= \frac{\sigma + \alpha x_{M-1}}{\alpha + \beta + \kappa}, \\
r(t) &= r = \alpha x_M.
\end{aligned} \quad (15)$$

Thus, a constant influx to the environmental nodes can be accounted for by adding a constant to the noise term. This implies that the only nontrivial solution ($\alpha \neq 0$) for the noise term is given by $\sigma = \kappa = 0$. Consequently, for this parameter setting the environment is mass conserving.

### A. Mapping onto a random walk

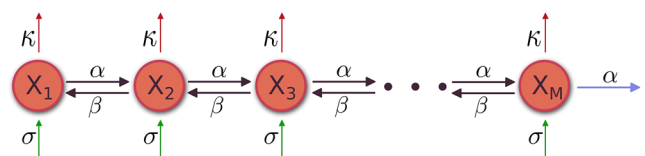Before calculating the memory function, we will first show that the considered environment can be mapped onto a



**FIG. 1**. Linear chain topology with mono-molecular interactions as considered as generic environment. The rates $\alpha$ and $\beta$ denote the transition rates between the different species $X_i$. $\sigma$ and $\kappa$ are the rates of constant in- and linear outflux, respectively. The blue arrow indicates the coupling to the subnetwork.

random-walk problem. Subsequently, we obtain a return-time distribution of the random walk problem, which is identical to the memory function of the chosen environment.

We explicitly model the interactions between the subnetwork and the environment. The subnetwork couples to the environment at some arbitrary position $x$, and the environmental chain couples back to the subnetwork at some other position(s) $y_{1/2}$. Since in this case the coupling from the subnetwork to the environment consists only of one single reaction, $\mathbf{A}^{ES}$ will only have one non-zero element. Furthermore, we note that one key feature of a system of monomolecular reactions is that there is no interaction between different molecules in the network, i.e., the reaction rate of a molecule is independent from that of all the others. Hence, inside the environment the mean evolution equation of the system's state is equal to its probability evolution equation. To obtain the general mean equation for the system with N particles, we need to scale up,

$$\langle \boldsymbol{x} \rangle_t = N \boldsymbol{p}_t.$$

Since, additionally, for a linear system the mean equation is identical to the deterministic evolution equation—which we use in our calculations—it is possible to describe the dynamics of our environmental system by

$$\boldsymbol{\mu}^E = N \boldsymbol{p}^E,$$

where N is the number of particles in the system. Additionally, for networks with mono-molecular interactions it was shown that an initial multinomial distribution $M(\boldsymbol{x}, \boldsymbol{p}(0))$ stays multinomial over time.[24] The parameter $\boldsymbol{p}$ of this distribution evolves according to the rate equation

$$p(t) = e^{\mathbf{B}t} p(0), \tag{16}$$

where $\mathbf{B}$ is the deterministic transition rate of the network. In a network with an absorbing state $x_{abs}$ and corresponding $p_{abs}$, the probability to be already absorbed at time $t$ is then given by

$$P_{abs}(t) = 1 - M(x_{abs} = 0, \boldsymbol{p}(t)) = p_{abs}(t).$$

Hence, the probability to get absorbed at time $t$ is

$$\frac{dP_{abs}}{dt} = \frac{dp_{abs}}{dt} = \mathbf{B}^{AN} \boldsymbol{p}(t), \tag{17}$$

where $\mathbf{B}^{AN}$ contains the transition rates from the network into the absorbing state.

Using this feature of systems with mono-molecular interactions, we can reinterpret the memory kernel in (14). The second part $\mathbf{A}^{ES} \boldsymbol{\mu}^S(\tau) d\tau =: \boldsymbol{p}^{ES}(\tau)$ in (14) is the concentration that enters the environment during the time interval $d\tau$ coming from the subnetwork. This initial distribution of one species can be interpreted as a multinomial distribution with $\boldsymbol{p} = \boldsymbol{e}_i$. In the first part of (14), the matrix $\mathbf{A}^{SE}$ denotes [similarly to $\mathbf{B}^{AN}$ in (17)] the rate to be absorbed in the subnetwork. The

term $e^{\mathbf{A}^{EE}(t-\tau)}$ is the deterministic solution of (16), where $\mathbf{A}^{EE}$ is the deterministic rate matrix of the environment only. Hence, $\mathbf{A}^{SE} e^{\mathbf{A}^{EE}(t-\tau)} =: \boldsymbol{p}^{SE}(t-\tau|\tau)$ denotes the rate to leave the environment after a time $t - \tau$ by entering the subnetwork again, given that the environment was entered at time $\tau$. Naturally, a particle that decays inside the environment will no longer be able to couple back to the subnetwork. Since there is no interaction between the single particles, this directly suggests how linear decay enters the memory terms

$$\boldsymbol{p}^{SE}(t - \tau \mid \tau) = \mathbf{A}^{SE} e^{\bar{\mathbf{A}}^{EE}(t-\tau)} e^{-\kappa(t-\tau)}, \tag{18}$$

where $\bar{\mathbf{A}}^{EE}$ denotes the environmental interactions without decay terms. We now reinterpret each environmental state (denoted by $X_i$ in Fig. 1) as a possible position of a random walker and take each (linear) reaction rate as the corresponding transition probability between the states. As a consequence, to obtain $\mathbf{A}^{SE} e^{\mathbf{A}^{EE}(t-\tau)} =: \boldsymbol{p}^{SE}(t-\tau|\tau)$, we can use a formalism used to analyze random walks with absorbing boundaries to account for the coupling to the subnetwork. Independently from us, Stephan and Stephan[25] developed a complementary approach to tackle similar linear environments.

## B. Absorption probability for continuous-time random walk

We will now derive a general expression $\boldsymbol{p}^{SE}(t - \tau|\tau)$ for any type of a linear random walk with absorbing boundaries. The problem at hand is a general question of first-return times of a one-dimensional random walk, before coming to more specific situations in Secs. III C 1–III C 3. In order to simplify the notation, we will enumerate the states in the chain with $\omega = 1, \ldots, \Omega$. As we will show later, it is the most general approach to start with a description that has two absorbing states, namely, the states $\omega = 0$ and $\omega = \Omega$. We use the backward master equation to describe the probability $P_{\omega,\chi}(t - \tau)$ of a state starting at $\omega$ to have reached the absorbing state $\chi$ at time $t - \tau$,

$$\frac{dP_{\omega,\chi}(t)}{dt} = -(\alpha + \beta)P_{\omega,\chi}(t) + \alpha P_{\omega+1,\chi}(t) + \beta P_{\omega-1,\chi}(t),$$

where $\alpha$ and $\beta$ are the transition probabilities of one step to the right and left, respectively. A problem like this was solved by Heathcote and Moyal,[26] where they also assumed a chain with absorbing barriers at $\omega = 0$ and $\omega = \Omega$. They were able to derive the following equations for the probability to be absorbed at 0 or $\Omega$ at time $t$:

$$\mathscr{P}_{\omega,0}(t) = \nu^{-\omega} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\mu\tau} \Big( (2j\Omega + \omega) I_{2j\Omega+\omega}$$
$$- (2(j+1)\Omega - \omega) I_{2(j+1)\Omega-\omega} \Big) d\tau,$$

$$\mathscr{P}_{\omega,\Omega}(t) = \nu^{\Omega-\omega} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\mu\tau} \Big( ((2j+1)\Omega - \omega) I_{((2j+1)\Omega-\omega)}$$
$$- ((2j+1)\Omega + \omega) I_{((2j+1)\Omega+\omega)} \Big) d\tau,$$

where $\mu = \alpha + \beta$, $\nu = \sqrt{\frac{\alpha}{\beta}}$, and $I_x$ is the modified Bessel function of the first kind with argument $2\tau\sqrt{\alpha\beta}$. Since both final states 0 and $\Omega$ are absorbing, the probabilities are of cumulative nature and are hence denoted by $\mathscr{P}$ instead of the previously used $P$. We are now interested in the probability to be absorbed exactly after a time-period of $\tau$ which is formally equal to the rate of leaving the environment and thus need the time derivative of these equations

$$
\begin{aligned}
P_{\omega,0}(t) &= \nu^{-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \big( (2j\Omega + \omega) I_{2j\Omega+\omega} \\
&\quad - (2(j+1)\Omega - \omega) I_{2(j+1)\Omega-\omega} \big), \\
P_{\omega,\Omega}(t) &= \nu^{\Omega-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \big( ((2j+1)\Omega - \omega) I_{((2j+1)\Omega-\omega)} \\
&\quad - ((2j+1)\Omega + \omega) I_{((2j+1)\Omega+\omega)} \big).
\end{aligned}
\tag{19}
$$

Since the environment is finite, we have

$$
\int_0^{\infty} dt \, (P_{\omega,0}(t) + P_{\omega,\Omega}(t)) = 1,
$$

which means that the mass is conserved.

In the following, we will show how we can use the random-walk model to describe different types of generic environments and derive the corresponding memory kernels using the equations that we just derived.

## C. Examples of environments and memory kernels

### 1. One single species

To give some intuition into the effects of environmental reactions, let us start with one of the simplest examples of a possible environment, namely, that of just one species coupled to the subnetwork with rates $\alpha$ and $\beta$ [Fig. 3(a)]. Examples of this type of context effect are, e.g., the reversible binding of one subnetwork species to an environmental site. In this case, the only non-zero elements of the coupling matrices are $A^{SE} = -A^{EE} = \beta$ and $A^{ES} = \alpha$, and the only non-zero element of the memory-kernel (14) simplifies to

$$
M(\tau) = \beta e^{-\beta\tau} \alpha. \tag{20}
$$

This is the expected exponential probability distribution of any quantity decaying with a linear rate.

### 2. Chain with two connections to the subnetwork

We now consider the most general configuration where the subnetwork couples to an environmental chain, as depicted in Figs. 2(a) and 2(b). The chain couples back to the subnet and can hence be left by the random walker at both ends. Examples for this type of environment can, e.g., be found in promiscuous conformational changes of proteins. Here, a protein is only functional in its native state (contained in the subnetwork's description). Due to structural variations like fold transitions, the protein can transform into other dysfunctional states that form a loop-like reaction topology (described by the environment).[22] Thus, we can directly use the sum of Eq. (19) that describes a finite chain with absorbing states at
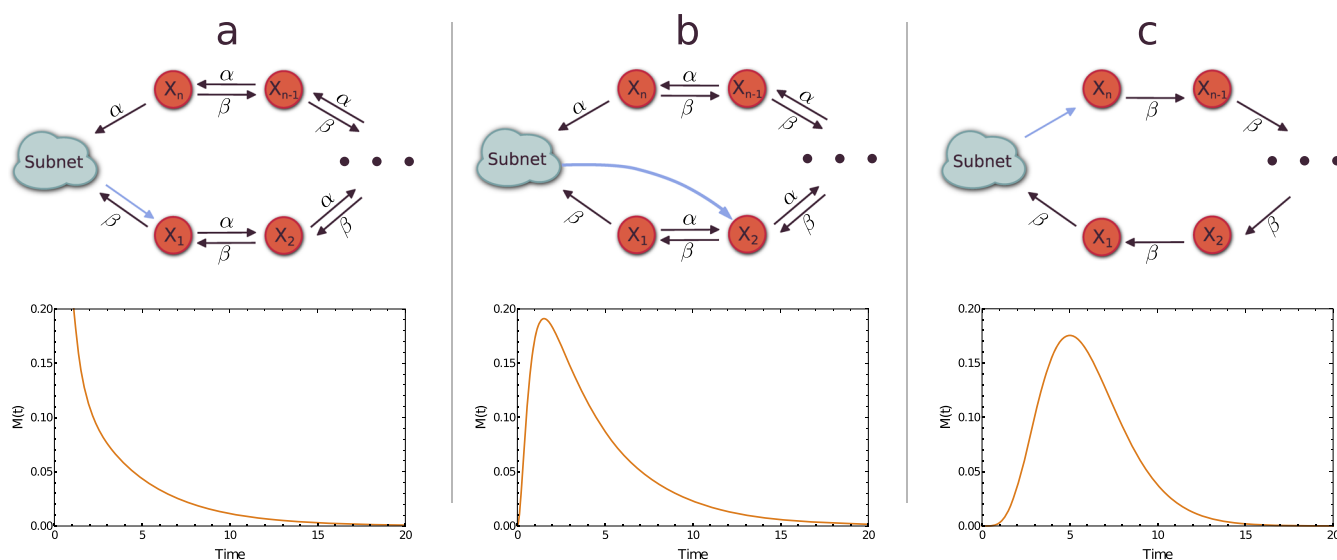


**FIG. 2**. Exemplary sketch of the topology (top) and the corresponding shape of the memory function (bottom) of circular environments as described in Sec. III C 2. [(a) and (b)] Bi-directional linear chain with coupling to the environment at (a) the first position and (b) some intermediate position [see (21)] and (c) directed linear chain [see (22)]. Parameters are $\alpha = \beta = 1$ and $\Omega = 6$. Additionally for (b) $\omega = 3$. The three dots denote an arbitrary long continuation of the chain.

both ends to obtain the memory kernel of such a system

$$M_{\text{biChain}}(t) = P_{\omega,0}(t) + P_{\omega,\Omega}(t). \quad (21)$$

In the given example of the promiscuous conformational changes of proteins, the coupling from the subnetwork to the environment is at the first position of the chain [Fig. 2(a)], hence $\omega = 1$, but in general there is no constraint on the coupling from the subnetwork to the environment [Fig. 2(b)].

Enzymatic reaction cycles are often highly directional, which means in our case that we can set $\lim_{\alpha \leftarrow 0}$ [Fig. 2(c)]. Thus, there is no longer any absorption at position $\omega = \Omega$, hence $P_{\omega,\Omega}(t) = 0$ and we obtain

$$M(t) = \lim_{\alpha \to 0} P_{\omega,0}(t) = \frac{\beta^\omega e^{-t\beta} t^{\omega-1}}{\Gamma(\omega)}. \quad (22)$$

As expected, this resembles the well-known Erlang distribution. For a chain length of $\omega = 1$, we obtain the same result as in (20).

### 3. Linear chain with one connection to the subnetwork

Letting $\Omega \to \infty$ in (21), we minimize the influence of the second absorbing barrier and the probability to be absorbed at $\omega = \Omega$ goes to zero; hence, we arrive at the description of some environment with only one coupling to the subnetwork. In this case, the probability to be absorbed at $\omega = 0$ at time $t$ simplifies to[26]

$$\lim_{\Omega \to \infty} P_{\omega,0}(t) = \lim_{\Omega \to \infty} \nu^{-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \big((2j\Omega + \omega)I_{2j\Omega+\omega}$$

$$- (2(j+1)\Omega - \omega)I_{2(j+1)\Omega-\omega}\big)$$

$$= \omega \nu^{-\omega} t^{-1} e^{-\mu t} I_\omega. \quad (23)$$
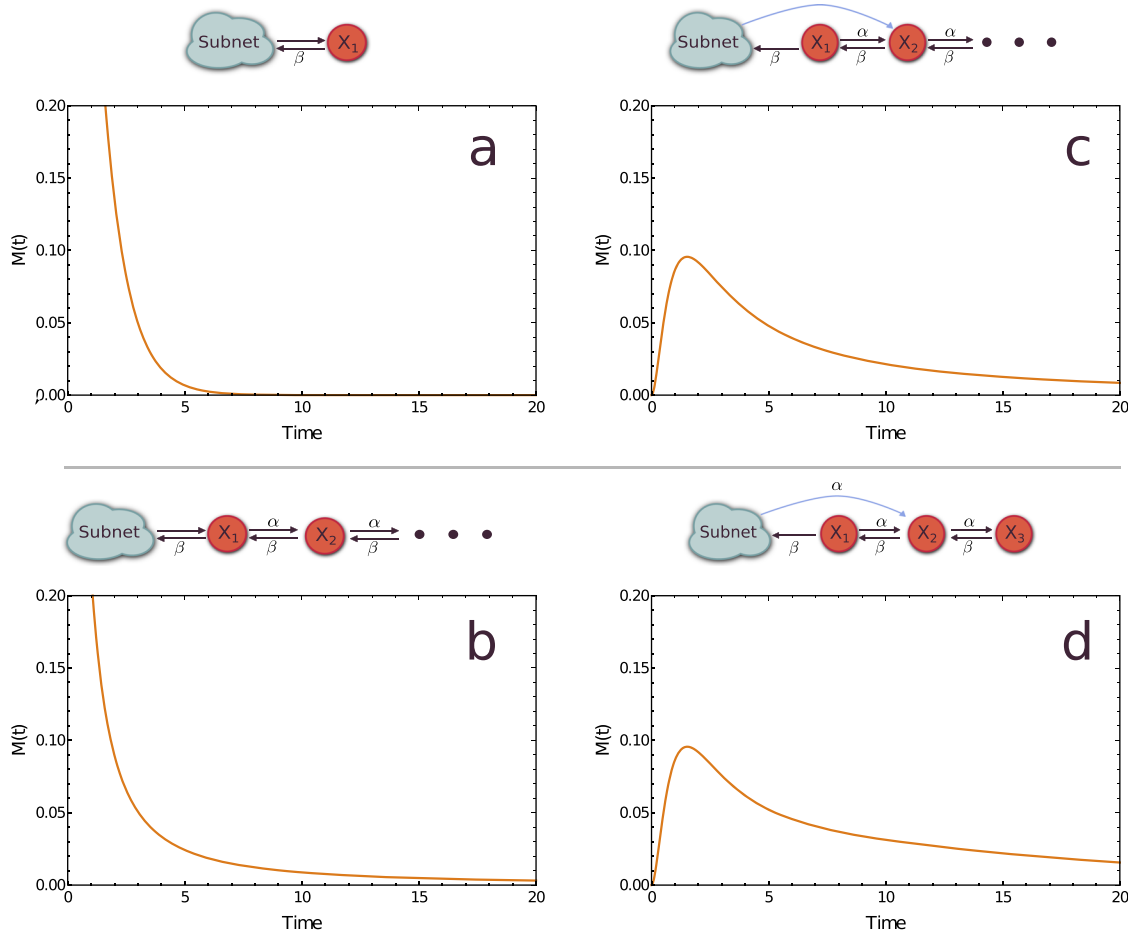


**FIG. 3**. Sketch of the topology (top) and the corresponding shape of the memory function (bottom) of linear environments with only one coupling to the environment as described in Sec. III C 3. (a) Environment with single species [see (20)], (b) infinite chain of interconnected species, where the coupling between the subnet and the environment is at the same position [see (23) with $\omega = 1$], (c) infinite chain of interconnected species, where the coupling between the subnet and the environment is at different positions [see (23) with $\omega > 1$], and (d) finite chain of interconnected species [see (24)]. Parameters are $\alpha = \beta = 1$. Additionally for (c) $\omega = 3$ and for (d) $\Omega = 6$, $\omega = 3$.
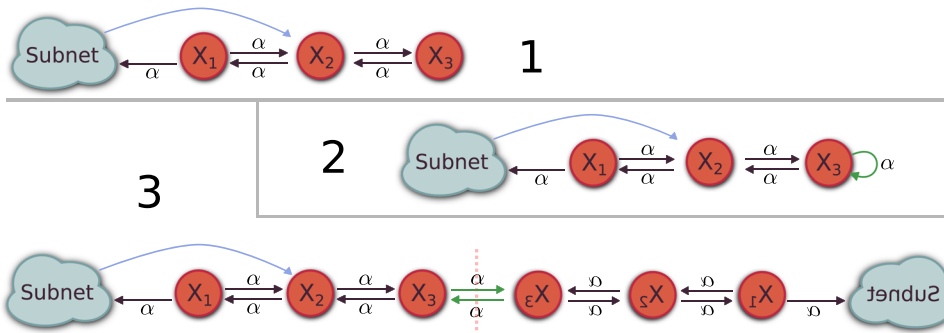
**FIG. 4**. Sketch for the equivalent transformation of a finite chain to a two-ended chain. (1) The initial finite chain we want to describe is extended by a self-referring loop reaction (2, green). This added reaction can be considered as a transition reaction between two equivalent finite chains that now form a two-ended network topology (3). The "mirror" axes are depicted as a pink dashed line.

In comparison to the first example of a single-species environment [Fig. 3(a)], in a longer chain more than one species takes part in buffering species from the subnetwork [Fig. 3(b)]. Hence, the memory function decays slower.

Similarly as for the double ended chain, a coupling to the environment at some intermediate position will lead to non-monotonous memory functions [Fig. 3(c)].

While infinite reaction chains might be a good approximation for long reaction cascades, chains with a small number of species are in general more relevant. Therefore, in the following, we consider reaction chains of finite length with one absorbing state at $\omega = 0$ and one reflecting state at $\omega = \Omega$ [e.g., see Fig. 3(d)], as depicted in Fig. 3(d) with $\alpha = \beta$.

At each position in the chain, a particle can move to the left and to the right with a summed propensity of $\alpha + \beta$. The last species (e.g., $X_3$ in Fig. 4) can only move to the left with summed propensity $\beta$. Nevertheless, without changing the dynamics in the system we can add a self-referring loop with rate $\alpha$ to $X_3$. This self-referring loop can also be considered as a connection between two different $X_3$ on two different, but interconnected finite chains (see Fig. 4). Hence, the dynamics of a particle on a chain with reflecting boundary condition is equivalent to the movement of a particle on a chain of length $2\Omega + 1$, where the absorption rate of the particle starting on position $\omega$ is given by

$$P_{\text{finiteChain}\Omega}(t) = \hat{P}_{\omega,0}(t) + \hat{P}_{\omega,2\Omega+1}(t), \tag{24}$$

where $\hat{P}$ denotes [just like in Eq. (19)] the rate of leaving the environment, but now with chain length $2\Omega + 1$.

## IV. APPLICATIONS

### A. Example 1: The effect of retroactivity on the repressilator

#### 1. Retroactivity

Del Vecchio *et al.* described a general engineering issue that arises as soon as different subsystems (modules) are interconnected to a large device.[6] They defined *retroactivity* as a signal that travels back from a downstream

module to an upstream module and derived a formula to quantify the effect of this retroactivity. A basic assumption of their derivation is the timescale separation between a fast downstream and a slow upstream module. The projection operator framework provides a more general method to analyze the impact of retroactivity on the performance of a genetic module, without relying on timescale separation. Having obtained memory kernels for different prototypic environments, we now give some examples of how context effects can drastically change the behavior of genetic devices.
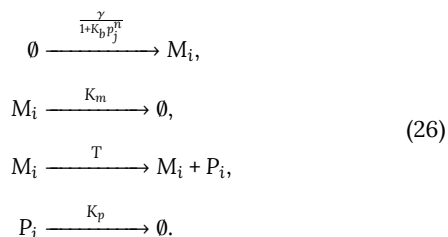
Del Vecchio *et al.* demonstrated the effect of retroactivity on an oscillating input protein signal $k(t)$ that is coupled to a downstream promoter binding region. In insulation, the system shows the desired oscillation, whereas, as soon as interconnected to the fast promoter, the oscillation is damped. In their study, $k(t)$ was used as a generic output signal of an upstream oscillating device. In the general case, the complete functioning of the oscillator will be deteriorated by the coupling to a downstream module. Using the example of the famous repressilator, we will explicitly model the system of interest in terms of reactions that result in an oscillating signal, and we will investigate the effect of retroactivity.

#### 2. Repressilator

The repressilator is a synthetic genetic regulatory network of three genes and the corresponding three proteins, where each protein represses another gene in a cyclic fashion. The resulting interaction can be described by the following equations:

$$\frac{dm_i}{dt} = -K_m m_i + \frac{\gamma}{1 + K_b p_j^n},$$
$$\frac{dp_i}{dt} = -K_p p_i + T m_i, \tag{25}$$

where $m_i$ denotes the concentration of the three different mRNAs and the $p_i$ of the proteins. The degradation rates of mRNA and proteins are given by $K_m$ and $K_p$, respectively. $T$ is the translation rate, $\gamma$, $n$, and $K_b$ are the Hill parameters. This corresponds to the following reaction scheme:

$$\emptyset \xrightarrow{\frac{\gamma}{1+K_b p_j^n}} M_i,$$

$$M_i \xrightarrow{K_m} \emptyset,$$

$$M_i \xrightarrow{T} M_i + P_i, \qquad (26)$$

$$P_i \xrightarrow{K_p} \emptyset.$$

Using the parameter set given by Elowitz and Leibler,[27] in the deterministic, as well as in the stochastic simulation, one observes a uniform oscillation in the concentrations of the different proteins [see Fig. 5 (left)].

### 3. Retroactivity due to promoter binding.

Using the repressilator, we can now build the system used by Del Veccio *et al.* where we, however, explicitly model the oscillating upstream module. One of the oscillating proteins of the repressilator is supposed to drive a downstream module by reversibly binding to its promoter. To account for this binding, we add to the reaction equations (26) a reversible binding reaction for protein A,

$$P_A \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C,$$
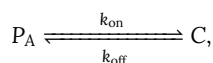
where C is the complex protein promoter. Using the results from (20), this leads to a modified set of differential equations

$$\dot{m}_i = -K_m m_i + \frac{\gamma}{1 + K_b p_j^n},$$

$$\dot{p}_i = -K_p p_i + T m_i + \delta_{Ai} \left( \int_0^t k_{off} e^{-k_{off}(t-\tau)} k_{on} p_i(\tau) d\tau - p_i k_{on} \right). \qquad (27)$$

For the isolated repressilator, one expects a driving frequency of the promoter analogous to the one depicted in Fig. 5 (left). Nevertheless—as shown by solving (27) as well as by stochastic simulations [Fig. 5 (right)]—due to the reversible binding of one of the proteins the oscillation frequency of the protein concentration is vastly reduced. Additionally, the amplitudes of the protein concentrations are changed due to the coupling to the promoter. This simple example highlights the need to account for context effects when designing synthetic genetic modules, especially, when time-scale separation is not possible.

### B. Example 2: Context effects due to promiscuity

We now want to proceed to a more sophisticated example of a rather counterintuitive result of context effects. We hereby build on a study by Bratsun *et al.*[28] Based on the observation that in gene regulatory networks there is often a vast separation of time scales between different types of reactions, the authors presented different gene regulation modules that show oscillations induced by delayed reactions. In the following, we modify their example of a "negative feedback with dimerization" by adding a constant inflow of X. Additionally, we motivate how the context of downstream modules can cause a delay.

A common function of synthetic genetic modules is the activation and repression of downstream genes. This is often achieved in two steps, where in the first step a specific monomer has to dimerize before it can bind to a downstream promoter. Denoting the monomer with X, the dimer with $X_2$, and the activated and inactivated promoter with D and $D^*$, respectively, the following reactions describe a repression of the promoter D by $X_2$:
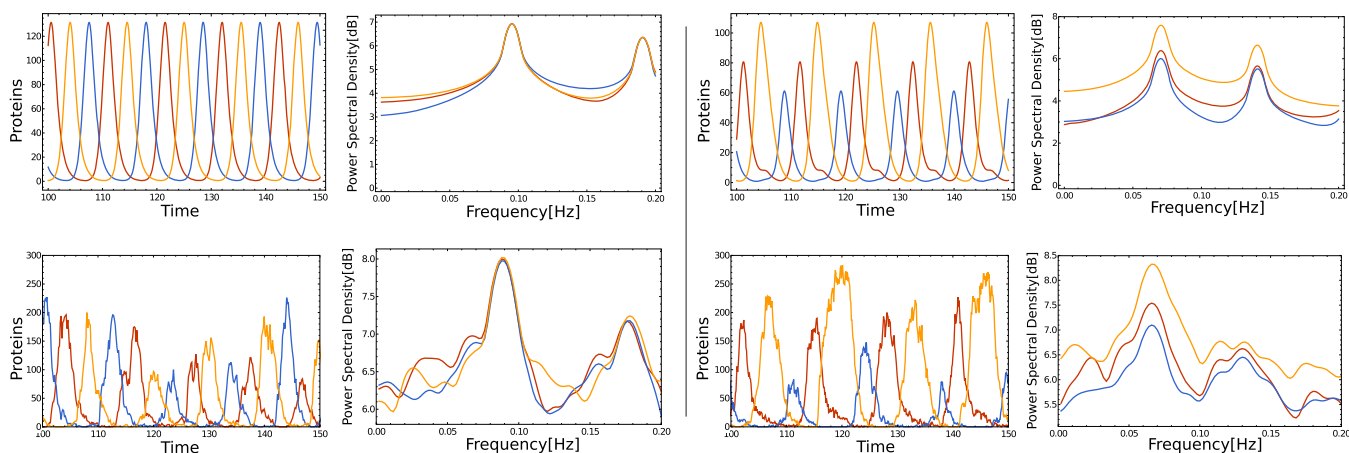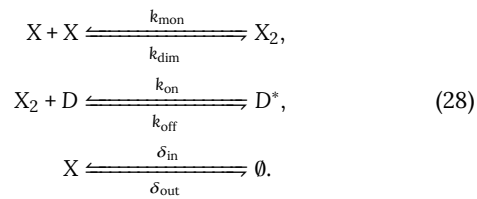


**FIG. 5**. Comparison of deterministic solution (top) and stochastic simulation (bottom) of the described repressilator system, where red, blue, and yellow denote $P_A$, $P_B$, and $P_C$, respectively. (left) The unmodified repressilator as given by (25). (right) The repressilator with one protein coupled to an environment as described in (27). Parameters taken from Elowitz and Leibler[27] are $K_p = T = 5$, $K_m = 1$, $n = 2.1$, $K_b = 1$, and $\gamma = 250$. Additionally parameters (right) taken from Ref. 6 are $k_{on} = 1000$ and $k_{off} = 100$.

$$X + X \underset{k_{\text{dim}}}{\overset{k_{\text{mon}}}{\rightleftharpoons}} X_2,$$

$$X_2 + D \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} D^*, \tag{28}$$

$$X \underset{\delta_{\text{out}}}{\overset{\delta_{\text{in}}}{\rightleftharpoons}} \emptyset.$$

Here, the relevant output parameter is the probability of an activated/inactivated promoter. The corresponding ODEs read

$$\dot{x}(t) = \delta_{\text{in}} - \delta_{\text{out}}\, x(t) - 2k_{\text{dim}}x(t)^2 + 2k_{\text{mon}}x_2(t),$$
$$\dot{x}_2(t) = -k_{\text{on}}d(t)x_2(t) + k_{\text{off}}d^*(t) + k_{\text{dim}}x(t)^2 - k_{\text{mon}}x_2(t),$$
$$\dot{d}(t) = -k_{\text{off}}d(t)x_2(t) + k_{\text{on}}d^*(t), \tag{29}$$
$$\dot{d}^*(t) = k_{\text{off}}d(t)x_2(t) - k_{\text{on}}d^*(t).$$

Analyzed in isolation, this results in a constant probability for an active promoter that depends on the chosen parameters

$$P(d) = \frac{\delta_{\text{out}}^2\, k_{\text{on}}\, k_{\text{mon}}}{\delta_{\text{out}}^2\, k_{\text{on}}\, k_{\text{mon}} + k_{\text{dim}}\, k_{\text{off}}}. \tag{30}$$

An example simulation as well as a numerical solution of the ODE system is shown in Fig. 6(a).

In a synthetic genetic network, the activated promoter will be used to trigger the operation of downstream modules. Let us assume that there is a downstream source of protein $\hat{X}$ that has—due to, e.g., promiscuity—properties similar to the X protein. To account for this, and assuming that expression of $\hat{X}$

occurs soon after the activation of the promoter $D$, one would add the following reaction to (28),

$$D \xrightarrow{\gamma} D + X \tag{31}$$
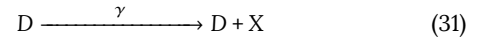
and likewise extend the first ODE in (29) that describes the time evolution of X,

$$\dot{x}(t) = \delta_{\text{in}} - \delta_{\text{out}}\, x(t) - 2k_{\text{dim}}x(t)^2 + 2k_{\text{mon}}x_2(t) + \gamma d(t).$$

The resulting behavior depicted in Fig. 6(b) essentially means that the feedback due to the interaction between downstream and upstream modules leads to a shift in the probability for an activated promoter. However, when we take into account the time delay between activation of $D$ and expression of $\hat{X}$, this result becomes modified. Such a time delay becomes significant if the number of modules between the dimerization module and the one expressing $\hat{X}$ is large or if there are some slow processes between the activation of $D$ and the expression of $\hat{X}$. We account for this by writing

$$D \xrightarrow{\gamma} D + Z_1,$$
$$Z_1 \xrightarrow{\gamma} Z_2,$$
$$Z_2 \xrightarrow{\gamma} Z_3,$$
$$\vdots$$
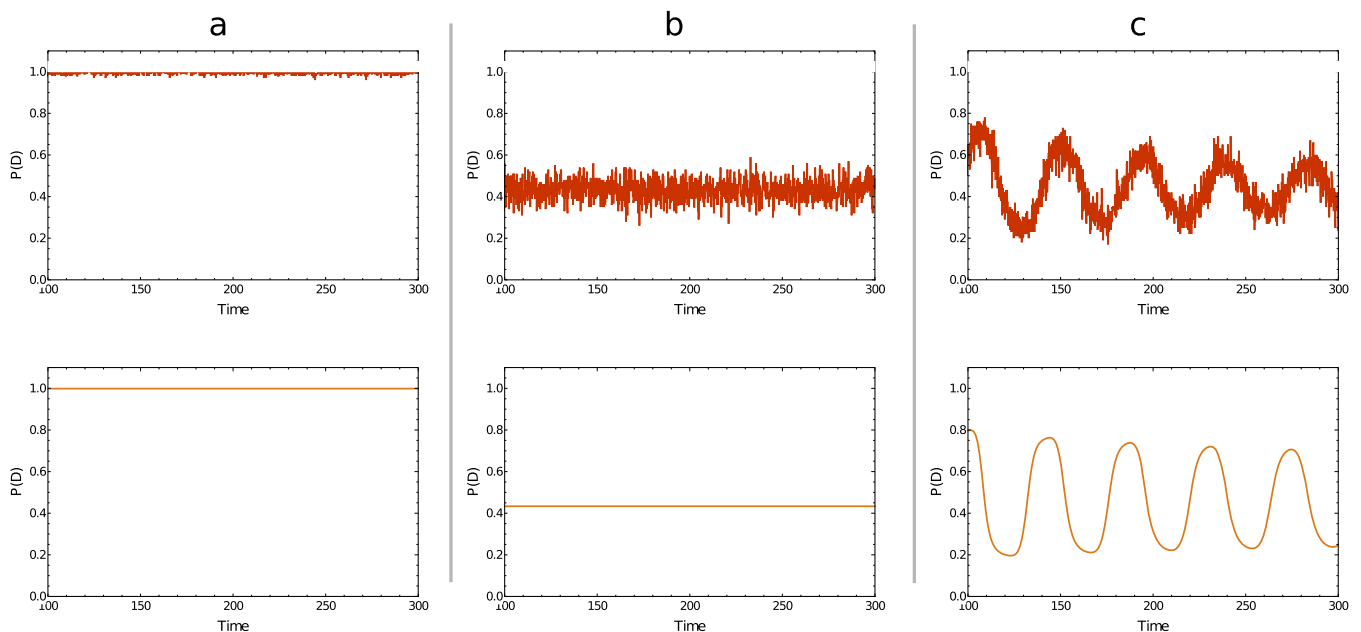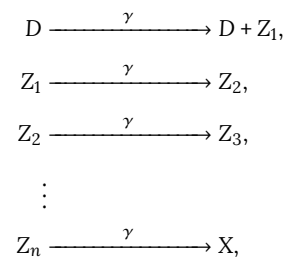$$Z_n \xrightarrow{\gamma} X,$$



FIG. 6. Comparison of deterministic solution (bottom) and stochastic simulation (averaged over 100 runs) (top) for the probability of an active promoter for the gene-regulation module described in Example 2. (a) The unmodified module without feedback, (b) the module with immediate feedback, and (c) the module with delayed feedback. Parameters are $\delta_{\text{in}} = 2$, $\delta_{\text{out}} = 4$, $k_{\text{on}} = 1000$, $k_{\text{off}} = 100$, $k_{\text{mon}} = 1000$, $k_{\text{dim}} = 200$, $\gamma = 70$, and $\tau = 20$.

where $Z_i$ are the intermediate steps before $\hat{X}$ is expressed. The chain of $Z_i$ reactions forms an environment similar to the directed chain described under Subsection III C 3. There, we have shown that $M(t - \tau)$ is an Erlang distribution as given by (22). We thus need to include this memory effect in the equation for $\dot{x}$. We will do this by using a simplification: if we keep the mean $\frac{\omega}{\alpha}$ fixed and increase the size of the chain, the variance decreases with $\frac{1}{\omega}$ and we can simplify

$$\lim_{\omega \to \infty} M(t - \tau)\Big|_{\frac{\omega}{\alpha}=\text{const}} = \lim_{\omega \to \infty} \frac{\beta^\omega e^{-(t-\tau)\beta}(t-\tau)^{\omega-1}}{\Gamma(\omega)}\Bigg|_{\frac{\omega}{\alpha}=\text{const}}$$
$$= \delta\left(t - \tau - \frac{\omega}{\alpha}\right). \quad (32)$$

As an approximation, we can hence treat the environment as a memory function that has the form of a time-delayed delta distribution. Using this result, we are now in the position to give an approximate result for the effects of the environment on the probability for an activated promoter. Taking into account this simple form of the memory effect, the equation for $\dot{x}$ now takes the form

$$\dot{x}(t) = \delta_{\text{in}} - \delta_{\text{out}}\, x(t) - 2k_{\text{dim}}x(t)^2 + 2k_{\text{mon}}x_2(t) + \int_0^t M(t-\tau)d(\tau)d\tau$$
$$= \delta_{\text{in}} - \delta_{\text{out}}\, x(t) - 2k_{\text{dim}}x(t)^2 + 2k_{\text{mon}}x_2(t) + \gamma\delta(t-\tau),$$

where we set $M(t - \tau) = \delta(t - \tau)$ as given by (32). In Fig. 6(c), we compare the stochastic simulation and the approximate deterministic solution of the self-repression circuit. Due to the context effects, the probability for the promoter to be activated now oscillates in time. As we have shown, this effect can only be observed if the memory effect of the environment is considered.

## V. CONCLUSION

In this article, we tackled the problem of context effects of synthetic genetic networks that are embedded in a cellular environment. To this purpose, we exploited the projection operator formalism to obtain an effective description of the target network that includes the environmental network via memory terms. In contrast to Rubin *et al.* who applied the projection framework to the chemical Langevin equation, we applied it directly to the chemical master equation to obtain the exact Nakajima-Zwanzig equation. For linear networks, we showed that the marginal RRE coincides with the first moment equation corresponding to the marginal/projected CME. This was, to the best of our knowledge, not done before.

Subsequently, we analyzed the memory terms of possible linear environments. While previous studies often assumed specific knowledge about the environment, to meet the requirements of synthetic biology, our goal was to avoid assuming specific knowledge of the environmental structure. Therefore, we considered several prototypic environments and deduced the memory kernels induced by them. In particular, we focused on environmental conversion chains of different lengths and with connections to the

subnetwork. Such reaction chains are found in a lot of scenarios of cellular environments. By exploiting the mapping between conversion chains and one-dimensional random walks, we found analytical expressions for the corresponding memory kernels.

Using the example of the repressilator, we applied our results and illustrated how the contributions of the context can deteriorate the system's performance. By coupling one protein species to an environmental promoter, we observed a noteworthy change in the oscillation frequency. Additionally, we used a simple self-repression module to demonstrate how memory terms can lead to counterintuitive phenomena like oscillations that cannot be detected with conventional analysis strategies.

Chain-like environmental cellular reaction structures occur in many cases, and most delay distributions can be approximated with unimolecular stochastic reaction networks.[29] Nevertheless, a reduction to only linear reactions is always an oversimplification, even though it may be a good approximation close to a steady state. Our results pose as a starting point for further analysis of non-linear environmental structures and more generic environments in the context of synthetic biology.

## ACKNOWLEDGMENTS

## APPENDIX: MARGINAL SECOND MOMENT EQUATION FOR LINEAR NETWORKS

Subsequently, we provide the derivation of the marginal second-order moment equation for the subnetwork species of a *completely linear* reaction network. Similarly to the derivation of the marginal mean equation in the main text, we start to characterize the action of $T^\dagger$ on quadratic functions. In particular, we define

$$\psi(\boldsymbol{x}) := \boldsymbol{x} \otimes \boldsymbol{x},$$

where $\otimes$ denotes the Kronecker product and we are interested in

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi\rangle_t = \left\langle T^\dagger\psi\right\rangle_t,$$

which is the equation for the second moments of the full network

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi(\mathbf{X})\rangle_t \equiv \frac{\mathrm{d}}{\mathrm{d}t}\langle\mathbf{X}\otimes\mathbf{X}\rangle_t = \sum_j \left\langle\mathbf{X}h_j(\mathbf{X})\right\rangle_t \otimes \boldsymbol{\nu}_j$$
$$+ \sum_j \boldsymbol{\nu}_j \otimes \left\langle\mathbf{X}h_j(\mathbf{X})\right\rangle_t + \sum_j (\boldsymbol{\nu}_j \otimes \boldsymbol{\nu}_j)\, h_j(\langle\mathbf{X}\rangle_t), \quad (A1)$$

where we used the linearity of the network in the last summand. As the resulting moment equations are closed and linear, we do not follow the operator formalism but rather apply

the more direct approach of partly solving the joint system explicitly to obtain a reduced set of equations for the moments of the subnetwork. For this purpose, we express (A1) and the first order dynamics as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\langle \mathbf{X}\rangle_t \\ \langle \mathbf{X}\otimes\mathbf{X}\rangle_t\end{bmatrix} = \begin{pmatrix}\mathbf{A} & \mathbf{0} \\ \mathbf{B} & \mathbf{C}\end{pmatrix}\begin{bmatrix}\langle \mathbf{X}\rangle_t \\ \langle \mathbf{X}\otimes\mathbf{X}\rangle_t\end{bmatrix}, \tag{A2}$$

where the first line corresponds to the mean equations determined by the block matrix $\mathbf{A}$ given by the RRE. We proceed to characterize the matrices $\mathbf{B}$ and $\mathbf{C}$.

Therefore, we exploit linearity to write $h_j(\boldsymbol{x}) = (\nabla h_j)\boldsymbol{x}$, where all entries of the gradients of the propensities are zero except one that is equal to the rate constant of the corresponding reaction. Thus, with

$$\mathbf{B} := \left[\sum_j\left(\begin{bmatrix}\boldsymbol{\nu}_j^{\mathrm{S}} \\ \boldsymbol{\nu}_j^{\mathrm{E}}\end{bmatrix}\otimes\begin{bmatrix}\boldsymbol{\nu}_j^{\mathrm{S}} \\ \boldsymbol{\nu}_j^{\mathrm{E}}\end{bmatrix}\right)(\nabla h_j^{\mathrm{S}},\nabla h_j^{\mathrm{E}})\right],$$

we can then write the last term of (A1) as

$$\sum_j(\boldsymbol{\nu}_j\otimes\boldsymbol{\nu}_j)\,h_j(\langle\mathbf{X}\rangle_t) = \mathbf{B}\begin{bmatrix}\langle\mathbf{X}^{\mathrm{S}}\rangle_t \\ \langle\mathbf{X}^{\mathrm{E}}\rangle_t\end{bmatrix},$$

whereas for the first and second summand of (A1) captured by $\mathbf{C}$, we use the following identities. For column vectors $\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{h}$ of proper dimension, we have

$$\boldsymbol{x}\boldsymbol{x}^{\dagger}\boldsymbol{h}\otimes\boldsymbol{v} = (\mathbf{I}\otimes\boldsymbol{v}\otimes\boldsymbol{h}^{\dagger})\,\mathrm{vec}(\boldsymbol{x}\boldsymbol{x}^{\dagger}),$$

$$\boldsymbol{v}\otimes\boldsymbol{x}\boldsymbol{x}^{\dagger}\boldsymbol{h} = (\boldsymbol{v}\otimes\mathbf{I}\otimes\boldsymbol{h}^{\dagger})\,\mathrm{vec}(\boldsymbol{x}\boldsymbol{x}^{\dagger}).$$

With this, the remaining matrix $\mathbf{C}$ reads

$$\mathbf{C} = \sum_j\left\{\begin{pmatrix}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M\end{pmatrix}\otimes\begin{pmatrix}\boldsymbol{\nu}_j^{\mathrm{S}} \\ \boldsymbol{\nu}_j^{\mathrm{E}}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\nu}_j^{\mathrm{S}} \\ \boldsymbol{\nu}_j^{\mathrm{E}}\end{pmatrix}\otimes\begin{pmatrix}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M\end{pmatrix}\right\}$$
$$\otimes\left(\nabla h_j^{\mathrm{S}},\nabla h_j^{\mathrm{E}}\right).$$

Inserting $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in (A2) and interpreting those components of the differential equation involving environmental variables as non-autonomous equations, we can then solve them explicitly as a function of the variables involving only subnetwork variables. By inserting these expressions into the remaining differential equations involving only subnetwork variables, one obtains an autonomous integro-differential equation of the Mori-Zwanzig type describing the first and second order dynamics of the subnetwork. As the calculation is straightforward and the resulting equations are rather lengthy, we omit the explicit expressions.

## REFERENCES

[1] R. P. Shetty, D. Endy, and T. F. Knight, J. Biol. Eng. **2**, 5 (2008).

[2] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, Mol. Syst. Biol. **2**, 2006.0028 (2006).

[3] P. E. M. Purnick and R. Weiss, Nat. Rev. Mol. Cell Biol. **10**, 410 (2009).

[4] S. Cardinale and A. P. Arkin, Biotechnol. J. **7**, 856 (2012).

[5] J. Saez-Rodriguez, A. Kremling, and E. D. Gilles, Comput. Chem. Eng. **29**, 619 (2005).

[6] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag, Mol. Syst. Biol. **4**, 161 (2008).

[7] W. Liebermeister, U. Baur, and E. Klipp, FEBS J. **272**, 4034 (2005).

[8] K. J. Rubin, K. Lawler, P. Sollich, and T. Ng, J. Theor. Biol. **357**, 245 (2014).

[9] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, Nat. Methods **11**, 197 (2014).

[10] C. Zechner and H. Koeppl, PLoS Comput. Biol. **10**, e1003942 (2014).

[11] L. Bronstein and H. Koeppl, Phys. Rev. E **97**, 062147 (2018).

[12] Z. Cao and R. Grima, Nat. Commun. **9**, 3305 (2018).

[13] T. G. Kurtz, Ann. Probab. **8**, 682 (1980).

[14] R. Kühne and P. Reineker, Z. Phys. B: Condens. Matter **31**, 105 (1978).

[15] S. Nakajima, Prog. Theor. Phys. **20**, 948 (1958).

[16] R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[17] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[18] P. Thomas, R. Grima, and A. V. Straube, Phys. Rev. E **86**, 041110 (2012).

[19] D. Venturi and G. E. Karniadakis, Proc. R. Soc. A **470**, 20130754 (2014).

[20] J. H. Hofmeyr, H. Kacser, and K. J. van der Merwe, Eur. J. Biochem. **155**, 631 (1986).

[21] T. Schwander, L. Schada von Borzyskowski, S. Burgener, N. S. Cortina, and T. J. Erb, Science **354**, 900 (2016).

[22] N. Tokuriki and D. S. Tawfik, Science **324**, 203 (2009).

[23] K. Semrad, Biochem. Res. Int. **2011**, 532908.

[24] T. Jahnke and W. Huisinga, J. Math. Biol. **54**, 1 (2007).

[25] A. Stephan and H. Stephan, e-print arXiv:1804.02332 [math] (2018).

[26] C. R. Heathcote and J. E. Moyal, Biometrika **46**, 400 (1959).

[27] M. B. Elowitz and S. Leibler, Nature **403**, 335 (2000).

[28] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Proc. Natl. Acad. Sci. U. S. A. **102**, 14593 (2005).

[29] C. Briat and M. Khammash, e-print arXiv:1811.09188 [cs, math, q-bio] (2018).