Generalized fluctuation-dissipation theorem for non-Markovian reaction networks

Aimin Chen,¹ Huahai Qiu,^{2,*} Tianhai Tian¹,⁴ and Tianshou Zhou^{1,3,†}

¹School of Mathematics and Statistics, Henan University, Kaifeng 475004, China

²School of Mathematical and Physical Sciences, Wuhan Textile University, Wuhan 430200, People's Republic of China

³Key Laboratory of Computational Mathematics, Guangdong Province, and School of Mathematics,

Sun Yat-sen University, Guangzhou 510275, People's Republic of China

⁴School of Mathematics, Monash University, Melbourne 3800, Australia

(Received 18 November 2021; accepted 23 May 2022; published 10 June 2022)

Intracellular biochemical networks often display large fluctuations in the molecule numbers or the concentrations of reactive species, making molecular approaches necessary for system descriptions. For Markovian reaction networks, the fluctuation-dissipation theorem (FDT) has been well established and extensively used in fast evaluation of fluctuations in reactive species. For non-Markovian reaction networks, however, the similar FDT has not been established so far. Here, we present a generalized FDT (gFDT) for a large class of non-Markovian reaction networks where general intrinsic-event waiting-time distributions account for the effect of intrinsic noise and general stochastic reaction delays represent the impact of extrinsic noise from environmental perturbations. The starting point is a generalized chemical master equation (gCME), which describes the probabilistic behavior of an equivalent Markovian reaction network and identifies the structure of the original non-Markovian reaction network in terms of stoichiometries and effective transition rates (extensions of common reaction propensity functions). From this formulation follows directly the solution of the linear noise approximation of the stationary gCME for all the components in the non-Markovian reaction network. While the gFDT can quickly trace noisy sources in non-Markovian reaction networks, example analysis verifies its effectiveness.

DOI: 10.1103/PhysRevE.105.064409

I. INTRODUCTION

Intracellular processes in single cells are often depicted using biochemical reaction networks [1]. A main task of Systems Biology is to design biologically reasonable mathematical models that faithfully describe the dynamics of these networked systems. In many situations, intracellular reaction networks display large fluctuations in the molecular numbers or the concentrations of reactive species [2]. This property makes it imperative to analyze the random behavior of the biochemical reaction networks. Since Poisson statistics cannot be assumed to be valid for living systems, characterizing the stochastic properties of biochemical networks is challenging.

From the view of continuous time random walk (CTRW), biochemical reaction networks can be divided into two classes: Markovian and non-Markovian, where by "Markovian" we mean that reaction events happen in a memoryless manner, and by "non-Markovian" we mean that reaction events occur in a memory fashion. For the former class of networks, the reaction kinetics can be described by Poisson processes with constant rates, which are characterized by exponential distributions of inter-reaction waiting times [3,4]. This property enables the establishment of the chemical master equation (CME) describing the time evolution of the joint probability distribution for all the different components in the underlying Markovian system [4]. For the latter class of networks, however, inter-reaction waiting times in general follow nonexponential distributions such as Gamma distributions [5,6]. Nonexponential waiting times can create molecular memory between individual reaction events, leading to non-Markovian (or non-Poissonian) reaction kinetics. The increasing availability of time-resolved data on different kinds of interactions has also verified the extensive existence of molecular memory in intracellular reaction networks [7–14]. Recently, the so-called generalized CME (gCME) for a general reaction network with molecular memory has been established based on the CTRW theory [15-17].

In some cases, the CME or gCME can be analytically solved, often under simplified conditions, and some analytical methods have also been developed [18-21]. In most cases, however, it needs to be numerically solved using methods developed based on either direct integration [22-25] or Monte Carlo sampling [26-28]. The question with these numerical methods is that they are computationally expensive, and in particular, they become infeasible as the number of reactive species is large. This is because the number of equations in the CME or gCME grows exponentially with the number of reactive species, making stochastic analysis difficult even for empirical networks of moderate size [29,30].

As an approach to the modeling of intracellular reaction networks, the mesoscopic fluctuation-dissipation theorem (FDT) has been gaining increasing attention. In Markovian

^{*}qiuhuahai2006@163.com

[†]mcszhtsh@mail.sysu.edu.cn

reaction cases, FDT can be straightforward derived from the CME. In turn, the FDT can be used in fast characterization of the stochastic properties of complex biochemical reaction networks [4,31]. In non-Markovian reaction cases, however, the similar FDT has not been established although the corresponding gCME has been available [6,15]. This motivates the study of this paper. We will present a generalized FDT (gFDT) for a large class of reaction networks with general (exponential or nonexponential) reaction waiting-time distributions.

Since memory functions involved in the gCME are implicitly (in fact by Laplace transform) expressed by the given reaction waiting-time distributions [15], deriving the gFDT directly from the gCME is nontrivial. In order to overcome this difficulty, we first introduce an effective transition rate for every reaction in the network, which explicitly decodes the effect of molecular memory created due to nonexponential waiting-time distributions. Then, we construct an equivalent Markovian reaction network by taking the introduced effect transition rates as reaction property functions of the constructed Markovian reaction network. Importantly, the stationary CME for the constructed Markovian reaction network is exactly the same as that for the original non-Markovian reaction network. Finally, for the constructed Markovian reaction network, we make use of the extension of Kramers-Moyal expansions to derive the gFDT from the common stationary gCME, which can be used in the fast evaluation of fluctuations in the molecule numbers or the concentrations of reactive species in the original non-Markovian reaction network. The derive gFDT is apparently an extension of the classical linear noise approximation (LNA), and can provide insights into the role of molecular memory in controlling fluctuations in reactive species of interest in complex non-Markovian reaction networks. Two representative examples are analyzed to verify the prediction of the gFDT.

II. GENERAL THEORY

A. Reaction network formulism

Consider a general networked system of reactions with volume Ω and *N* different chemical components (denoted by X_1, \dots, X_N). Assume that these reactive species participate in *L* reactions in total:

$$\sum_{i=1}^{N} r_{ij} X_i \to \sum_{i=1}^{N} \tilde{r}_{ij} X_i, \quad j = 1, 2, \dots, L,$$
(1)

where r_{ij} and \tilde{r}_{ij} are stoichiometries. The concentrations of the system's components are summarized in column vector $\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$, whereas their molecular numbers in column vector $\mathbf{X} = \Omega \mathbf{x} = (X_1, \ldots, X_N)^{\mathrm{T}}$. The microscopic state of the system is defined by \mathbf{X} (or \mathbf{x}), and is changed by one of L reactions. Define $S_{ij} = \tilde{r}_{ij} - r_{ij}$ (net change, which may be negative, zero, or positive), where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, L$. These S_{ij} constitute an $N \times L$ stoichiometric matrix (\mathbf{S}) of the reaction network. In addition, denote by $\mathbf{S}_j = (S_{1j}, \ldots, S_{Nj})^{\mathrm{T}}$ the column vector of stoichiometric changes for reaction j.

Traditionally, describing the transition rates for reactions adopts reaction propensity functions if reaction events happen in a Markovian manner, e.g., $\sum_{i=1}^{N} r_{ij}X_i \xrightarrow{k_j} \sum_{i=1}^{N} \tilde{r}_{ij}X_i$, where k_j is a constant reaction rate. This is equivalent to the description of exponential waiting-time distributions, i.e., $\sum_{i=1}^{N} r_{ij} X_i \xrightarrow{\alpha_j(t, x; \Omega)} \sum_{i=1}^{N} \tilde{r}_{ij} X_i$, where $\alpha_j(t, x; \Omega) = h_j(x; \Omega)e^{-h_j(x; \Omega)t}$ in which $h_j(x; \Omega)$ is a reaction propensity function in the inter-reaction waiting-time distribution for reaction *j*. However, reaction events happen often in a non-Markovian manner with the reasons given in Refs. [6,7,15]. In the non-Markovian case, $\alpha_j(t, x; \Omega)$ may be a general (exponential or nonexponential) distribution, and will be assumed to a Gamma distribution: $\alpha_i^{(r)}(t; x) = [1/\Gamma(k_i)](\lambda_i(x))^{k_i}t^{k_i-1}e^{-\lambda_i(x)t}$, where $\lambda_i(x)$ is a regulatory function, $\Gamma(k_i)$ is a common Gamma function, and k_i is called memory index throughout this paper.

In order to better model the random occurrence of reaction events in terms of waiting times, we distinguish between intrinsic and extrinsic sources of noise. The intrinsic noise is referred to the inherent stochasticity of the reaction process proper [1,32,33]. A single event of reaction j is characterized by reaction waiting time $\tau_j^{(r)}$, whose probability density function (PDF), denoted by $\alpha_j^{(r)}(t, \mathbf{x}; \Omega)$ throughout this paper, depends, in general, on system state \mathbf{X} . The reaction event that actually occurs is the one whose waiting time is the minimum, that is, $\tau^{(r)} = \min\{\tau_j^{(r)}|1 \le j \le L\}$. Note that the expression $\prod_{j \ne k} [1 - \int_0^t \alpha_j^{(r)}(t', \mathbf{x}; \Omega)dt']$ or equivalently, $\prod_{j \ne k} \int_t^\infty \alpha_j^{(r)}(t, \mathbf{x}; \Omega)dt'$ represents the probability that the *k*th reaction does not happen. Thus, the joint distribution for reaction k to happen after a reaction time interval [t, t + dt] is given by $\kappa_k^{(r)}(t, \mathbf{x}; \Omega)dt$, where we define $\kappa_k^{(r)}(t, \mathbf{x}; \Omega) = \alpha_k^{(r)}(t, \mathbf{x}; \Omega) \prod_{j \ne k} \int_t^{+\infty} \alpha_j^{(r)}(t', \mathbf{x}; \Omega)dt'$.

In contrast to intrinsic noise, extrinsic noise results from external fluctuations, e.g., the variability in the physical or chemical environment. Under transport-limited conditions, reaction delays arise from mass-transfer limitations due to the spatial sampling efficiency and fluctuation-induced segregation of reactants [34,35]. In the kinetic Monte Carlo spirit, these delays can influence all particles in the same way, independently of the system state. Therefore, in contrast to $\tau^{(r)}$, we introduce a delay time, $\tau^{(d)}$, which does not depend, directly, on the state but may depend on the current reaction waiting time, $\tau^{(r)}$. That is, for a given state X, the reaction waiting time takes the form $\tau = \tau^{(r)} + \tau^{(d)}(\tau^{(r)})$. Let $\tau_i^{(d)}$ be a stochastic reaction delay for reaction *i* due to environmental perturbations and assume that this delay follows a distribution, $\kappa_i^{(d)}(t)$. Then, the actual PDF for the waiting time of reaction i should be $\kappa_i(t, \boldsymbol{x}; \Omega) = (\kappa_i^{(r)} * \kappa_i^{(d)})(t, \boldsymbol{x}; \Omega)$, where * denotes convolution. Intuitively, the delay time can be viewed as a "preparation" time for the next reaction event to happen.

Regarding reaction delays, here we emphasize the following four points: (1) the delay time is external but the delay event is triggered by the reaction event; (2) the reaction delays generated by environmental heterogeneity or by the extrinsic noise mean that reaction events are postponed to happen in contrast to the case of no extrinsic noise; (3) the reaction delays in our model, which are implicit, are different from those in delay models in Ref. [36], which are explicit, where the former means that the extrinsic noise impacts the occurrence of each reaction, whereas the latter means that a delay only affects the rate (rather than the occurrence) of a particular reaction; and (4) the delay of each reaction in our model is random and is assumed to follow a distributions, whereas the delay is extracted from a given distribution and is essentially deterministic. In a word, the meaning and source of two kinds of delays are completely different.

Note that if reaction events happen in a Markovian (or memoryless) manner, then the intrinsic-event waitingtime distribution for reaction j, $\alpha_j^{(r)}(t, \mathbf{x}; \Omega)$, is exponential. Moreover, this distribution takes the form $\alpha_j^{(r)}(t, \mathbf{x}; \Omega) = h_j(\mathbf{x}; \Omega)e^{-h_j(\mathbf{x};\Omega)t}$, where $h_j(\mathbf{x}; \Omega)$ is the reaction propensity function for reaction j ($1 \le j \le L$). In general, reaction events happen in a non-Markovian (or memory) fashion, implying that reaction waiting times follow nonexponential distributions.

B. Memory functions and effective transition rates

According to Ref. [15], we know that the joint probability that the *j*th reaction happens and the reaction waiting time is

within time interval [t, t + dt] given by $\kappa_j(t, \boldsymbol{x}; \Omega)dt$, where $\kappa_j(t, \boldsymbol{x}; \Omega)$ is the PDF for the waiting time of reaction j. Note that for reaction j of the reaction network, we may introduce a memory function $M_j(t, \boldsymbol{x}; \Omega)$ [6,15], which is defined by its Laplace transform with variable $t: \overline{M}_j(s, \boldsymbol{x}; \Omega) = \frac{s\overline{\kappa}_j(s, \boldsymbol{x}; \Omega)}{1 - \sum_{k=1}^{L} \overline{\kappa}_k(s, \boldsymbol{x}; \Omega)}$, whose form seems bizarre but is typical of the CTRW key formulism [37]. Here, functions $\overline{M}_j(s, \boldsymbol{x}; \Omega)$ and $\overline{\kappa}_j(s, \boldsymbol{x}; \Omega)$ are the Laplace transforms of functions $M_j(t, \boldsymbol{x}; \Omega)$ and $\kappa_j(t, \boldsymbol{x}; \Omega)$, respectively, and $j = 1, 2, \dots, L$. We point out that in general, memory function $M_j(t, \boldsymbol{x}; \Omega)$ cannot be explicitly expressed by waiting-time distributions. In the following, the above bar always represents Laplace transform.

Interestingly, we find that the limit $\lim_{s\to 0} \overline{M}_j(s, \boldsymbol{x}; \Omega)$ always exists. Moreover, if this limit is denoted by $W_i(\boldsymbol{x}; \Omega)$, it can be expressed as (see Appendix A for derivation)

$$W_i(\boldsymbol{x};\Omega) = \frac{\int_0^\infty \kappa_j^{(r)}(t,\boldsymbol{x};\Omega)dt}{\sum_{j=1}^L \langle \tau_j^{(d)} \rangle \int_0^\infty \kappa_j^{(r)}(t,\boldsymbol{x};\Omega)dt + \sum_{j=1}^L \int_0^\infty t\kappa_j^{(r)}(t,\boldsymbol{x};\Omega)dt},$$
(2)

where $\langle \tau_i^{(d)} \rangle = \int_0^{+\infty} t \alpha_i^{(d)}(t) dt$ represents the mean delay for reaction *i*. Throughout this paper, $W_i(\mathbf{x}; \Omega)$ is called the effective transition rate for reaction *i*. Equation (2) reveals that $W_i(\mathbf{x}; \Omega)$ depends, only, on the delay means, independently of the forms of the delay distributions. This is an interesting result. We point out that effective transition rates are extensions of reaction propensity functions [6,15].

Next, we consider the limit of large Ω , and define

$$\boldsymbol{\varphi} = \lim_{\substack{\Omega \to \infty \\ X \to \infty}} \Omega^{-1} \boldsymbol{X} = \lim_{\Omega \to \infty} \boldsymbol{x},$$
$$W_j(\boldsymbol{\varphi}) = \lim_{\Omega \to \infty} W_j(\boldsymbol{x}; \Omega),$$
$$j = 1, 2, \dots, L$$
(3)

where $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_N)^T$ is the system's macroscopic state. Apparently, if all the mean reaction delays are equal and the common delay is denoted by τ_g , Eq. (2) then becomes

$$W_{j}(\boldsymbol{\varphi}) = \frac{\int_{0}^{+\infty} \alpha_{j}^{(r)}(t;\boldsymbol{\varphi}) \prod_{i\neq j} \left[\int_{t}^{\infty} \alpha_{i}^{(r)}(t';\boldsymbol{\varphi}) dt' \right] dt}{\tau_{g} + \int_{0}^{+\infty} \prod_{i=1}^{L} \left[\int_{t}^{\infty} \alpha_{i}^{(r)}(t';\boldsymbol{\varphi}) dt' \right] dt},$$

$$j = 1, 2, \dots, L,$$
(4)

where $\alpha_j^{(r)}(t, \boldsymbol{\varphi}) = \lim_{\Omega \to \infty} \alpha_j^{(r)}(t, \boldsymbol{x}; \Omega)$, and we have made use of the fact $\sum_{j=1}^L \int_0^\infty \kappa_j^{(r)}(t, \boldsymbol{x}; \Omega) dt = 1$.

C. Generalized chemical master equations

Let $P(\mathbf{x}, t)$ be the probability that the system is in state \mathbf{x} at time t. According to the update theory [3,4], we can derive the following gCME for the above reaction network in the sense

of Laplace transform:

$$s\bar{P}(\boldsymbol{x},s) - P(\boldsymbol{x},0)$$

= $\Omega \sum_{j=1}^{L} \left(\prod_{i=1}^{N} \mathbb{E}^{-S_{ij}} - \mathbb{I} \right) [\bar{M}_{j}(s,\boldsymbol{x};\Omega)\bar{P}(\boldsymbol{x},s)],$ (5)

where $\overline{P}(\mathbf{x}, s)$ is the Laplace transform of probability function $P(\mathbf{x}, t)$. In Eq. (4), \mathbb{I} is a unit operator, \mathbb{E} is a step operator, and \mathbb{E}^{-1} is its inverse. Note that $\mathbb{E}^{-S_{ij}}$ is also a step operator and its operation rule is when $\mathbb{E}^{-S_{ij}}$ acts on a function, $g(X_1, X_2, \ldots, X_N)$, it removes S_{ij} molecules from reactive species X_i in reaction j, i.e., $\mathbb{E}^{-S_{ij}}g(X_1, X_2, \ldots, X_N) = g(X_1, \ldots, X_i - S_{ij}, \ldots, X_N)$. Thus, for any two integers i and k $(1 \le i, k \le N)$, we have the operation rule: $\mathbb{E}^{-S_{ij}}\mathbb{E}^{-S_{kj}}g(X_1, X_2, \ldots, X_N) =$ $g(\ldots, X_i - S_{ij}, \ldots, X_k - S_{kj}, \cdots)$ for reaction j $(1 \le j \le L)$.

Recall the final value theorem [38]: for arbitrary function f(t), if the limit $\lim_{t\to\infty} f(t)$ exists and is finite, two limits $\lim_{t\to\infty} f(t)$ and $\lim_{s\to 0} s\bar{f}(s)$ [where $\bar{f}(s)$ is the Laplace transform of function f(t)] are equal, namely $\lim_{t\to\infty} f(t) = \lim_{s\to 0} s\bar{f}(s)$. For the above reaction network, if stationary distribution P(x) exists (in fact, numerical simulation has verified this point [6,37]), we have

$$P(\mathbf{x}) = \lim_{t \to \infty} P(\mathbf{x}, t) = \lim_{s \to 0} s\bar{P}(\mathbf{x}, s).$$
(6)

Multiplying s on both sides of Eq. (5) and using Eq. (3) and Eq. (6), we thus obtain

$$\sum_{j=1}^{L} \left(\prod_{i=1}^{N} \mathbb{E}^{-S_{ij}} - \mathbb{I} \right) [W_j(\boldsymbol{x}; \Omega) P(\boldsymbol{x})] = 0,$$
(7)

which is called the stationary gCME for the above reaction network with system size Ω . This equation lays a solid foundation for the further analysis of this paper.



FIG. 1. Schematic diagram for two reaction networks with the same topology and the same species. α -Type functions in the lefthand side diagram represent inter-reaction waiting-time distributions and are assumed to take the forms in this paper: $\alpha_i^{(r)}(t; \mathbf{x}) = [1/\Gamma(k_i)](\lambda_i(\mathbf{x}))^{k_i}t^{k_i-1}e^{-\lambda_i(\mathbf{x})t}$, where $\lambda_i(\mathbf{x})$ are regulatory functions depending on the system's microscopic state $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ with x_j representing the concentration or the molecular number of reactive species X_j , and shape parameter k_i of distribution $\alpha_i^{(r)}(t; \mathbf{x})$ is called memory index throughout this paper; W-type functions in the right-hand side diagram are effective transition rates [Eq. (2)]; and system size Ω is not explicitly considered or is omitted. The model to be studied also considers stochastic reaction delays (see the main text). The results in Ref. [6] imply that the stationary behaviors of two reaction networks are exactly the same although reaction events occur in different manners.

By comparing the structure of Eq. (5) and Eq. (7), we naturally construct a reaction network as shown in the left-hand side diagram of Fig. 1: it has the same network topology as the original reaction network in the right-hand side diagram of Fig. 1) and takes $W_j(\mathbf{x}; \Omega)$ as its reaction propensity functions. Clearly, the stationary CME for this constructed reaction network is nothing but Eq. (7), implying that the steady-state behavior of the constructed Markovian reaction network is exactly the same as that of the original non-Markovian reaction network. In fact, if we list the CME for the constructed reaction network in the sense of Laplace transform

$$sQ(\mathbf{x}, s) - Q(\mathbf{x}, 0)$$

= $\Omega \sum_{j=1}^{L} \left(\prod_{i=1}^{N} \mathbb{E}^{-S_{ij}} - \mathbb{I} \right) [W_j(\mathbf{x}; \Omega) \bar{Q}(\mathbf{x}, s)], \quad (8)$

the form of which is similar to that of Eq. (5), where $\bar{Q}(\mathbf{x}, s)$ is the Laplace transform of probability function $Q(\mathbf{x}, t)$ for the constructed network; then, the CME of Eq. (8) at steady state is Eq. (7) except for notation. This property will greatly simplify the analysis of non-Markovian reaction networks.

In the following, we will restrict our analysis to the constructed Markovian reaction network due to the equivalence of the stationary behaviors of the two reaction networks (in spite of some differences in dynamical behaviors at initial time points between the two reaction networks [39]).

D. Generalized fluctuation-dissipation theorem

For the reaction network constructed above, the relative fluctuations in components will eventually become insignificant if the system volume Ω is sufficiently large. In the limit of large Ω , the stochastic concentration vector \mathbf{x} will normally become deterministic, and the effective transition rates, $W_j(\mathbf{x}, \Omega)$, will be simplified to their macroscopic rate law counterparts, $W_j(\boldsymbol{\varphi})$. On the other hand, the LNA is often used in simplifying the analysis of reaction networks [3,4,40]. For the above non-Markovian reaction system, here we present a generalized fluctuation-dissipation theorem, which can be used in the fast evaluation of stochastic fluctuations in a reactive species of interest.

First, note that macroscopic concentrations at steady state satisfy the following N algebraic equations:

$$\sum_{j=1}^{L} S_{ij} W_j(\boldsymbol{\varphi}) = 0, \quad i = 1, 2, \dots, N,$$
(9)

which altogether determine the steady state (denoted by $\tilde{\varphi}$) of the deterministic system corresponding to the constructed Markovian reaction network. This steady state can be taken as the one of the original non-Markovian reaction system. We point out that these stationary equations can also be derived directly from Eq. (7).

Second, it is widely agreed that if the average copy number is *n*, the size of fluctuations around it must be \sqrt{n} (Poisson statistics). This means that if the average molecular number (*n*) of a certain species in an *Escherichia coli* cell is 100 or more (corresponding to a concentration >0.2 μ M [41]), the relative fluctuations ($\sqrt{n}/n = 1/\sqrt{n}$) would be around 10% or less. Accordingly, if the molecular number is large (>1000), the relative fluctuations would be only 3% and therefore insignificant. However, intracellular chemical reactions take place far from equilibrium. Therefore, Poisson statistics assumptions may not be valid, and the stochastic fluctuations around the average of *n* molecules may be much larger or much smaller than \sqrt{n} [42,43]. In general, we can write X = $\Omega x + \Omega^{1/2} \xi$, where ξ is a vector of random variables.

Now, we state the following theorem (see the proof details given in Appendix B of this paper), which is the main result of this paper:

The generalized fluctuation-dissipation theorem: Let $\tilde{\varphi}$ be the macroscopic stationary state of the networked system, and denote $A_{ij} = \sum_{k=1}^{L} S_{ik} \partial W_k(\tilde{\varphi}) / \partial \varphi_j$ and $[\mathbf{BB}^T]_{ij} = \sum_{k=1}^{L} S_{ik} S_{jk} W_k(\tilde{\varphi})$, where i, j = 1, 2, ..., N. Then, we have the Lyapunov matrix equation:

$$\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^{\mathrm{T}} + \Omega \mathbf{B}\mathbf{B}^{\mathrm{T}} = 0, \qquad (10)$$

where $\mathbf{C} = \langle (X - \langle X \rangle)(X - \langle X \rangle)^T \rangle \equiv \langle \Delta X \cdot \Delta X^T \rangle$ is a covariance matrix.

In Eq. (10), matrices **A** and **B** are all known, but matrix **C** is unknown. Note that these three matrices are all square ones of $N \times N$ orders, and in particular, the diagonal elements of matrix **C** represent variances. This theorem is apparently an extension of the LNA in the Markovian case and is therefore called the generalized LNA (gLNA). This LNA is apparently an extension of the common (i.e., in Markovian cases) LNA. This is because the effective transition rates [i.e., $W_k(\varphi)$, $1 \leq k \leq L$] are extensions of common reaction-propensity functions of Markovian reaction systems, and only when all waiting-time distributions are exponential or all the reaction rates are constants do our effective transition rates reduce to common reaction-propensity functions.

If the noise is quantified by the ratio of the variance over the squared mean, we can calculate the noise intensities for individual reactive species according to

$$\eta_{X_i} = \frac{C_{ii}}{\Omega^2 \tilde{\varphi}_i^2} = \frac{\Sigma_{ii}}{\Omega \tilde{\varphi}_i^2}, \quad i = 1, 2, \dots, N,$$
(11)

where Σ_{ii} are diagonal elements of matrix $\Sigma = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle \equiv \langle \Delta \mathbf{x} \cdot \Delta \mathbf{x}^T \rangle = \frac{C}{\Omega}$, and $\tilde{\varphi}_i$ is a component of vector $\tilde{\varphi}$. Similarly, we can give calculation formulae for the Fano factors of reactive species if the Fanbo factor is defined as the ratio of the variance over the mean. Note that $\Omega \tilde{\varphi}_i$ is the macroscopic steady-state molecular number for reactive species X_i . This theorem can provide a fast evaluation of stochastic fluctuations in individual reactive species in any reaction network.

III. EXAMPLE ANALYSIS

Here, we choose two simple yet representative examples for analysis, so that approximate solutions of the stationary gCME can be analytically derived. Then, we use these approximations to clarify the sources of fluctuations in each case and to trace the effects of different parameters on the stochastic properties of the systems.

A. Importance of molecular memory and stoichiometries: Analysis of a generalized birth-death process

As one of the most common and fundamental reaction modules, the birth-death process is a simple, natural, and formal framework for modeling many biological processes such as constitutive gene expression [6,39], population dynamics, speciation, and genome evolution [3,4,44,45]. Such a simple reaction module was previously studied under the Markovian hypothesis. Main results include that the stationary number of molecules for the reactive species follows a Poissonian distribution [4] and the corresponding Fano factor is equal to 1. These results well characterize the reaction kinetics of birth-death processes without molecular memory. As pointed out in the Introduction, however, the waiting times between reaction events in general follow nonexponential distributions, raising issues: how does the non-Markovianity affect reaction kinetics? And, does the non-Markovianity have extra effects in contrast to Markovianity? Addressing these issues is not nontrivial even for non-Markovian birth-death processes since non-Markovianity can induce additional dynamics [6].

The first example to be analyzed is a system that contains only one type of reactive species, X. The X molecules are synthesized and consumed by two different reactions with general waiting-time distributions:

$$\emptyset \xrightarrow{\alpha_1^{(r)}(t;\varphi)} mX, \quad nX \xrightarrow{\alpha_2^{(r)}(t;\varphi)} \emptyset, \tag{12}$$

where φ is the molar concentration of *X*, *m* and *n* are stoichiometries and two positive integers. The first reaction adds *n* molecules to *X* in a single event, and the second reaction removes *m* molecules from *X* in another single event. This example may describe many molecular systems, e.g., those of constitutive gene expression [6,39]. In fact, any reactive species will undergo both a birth process and a death process. Apparently, if *m* = 1 and *n* = 1, and if both $\alpha_1^{(r)}(t;\varphi)$ and $\alpha_2^{(r)}(t;\varphi)$ are exponential functions, this corresponds to the common birth-death process.

Below, we set $\alpha_1^{(r)}(t;\varphi) = \frac{t^{k_1-1}e^{-\mu t}}{\Gamma(k_1)}\mu^{k_1}$ and $\alpha_2^{(r)}(t;\varphi) = \frac{t^{k_2-1}e^{-\lambda(\varphi)t}}{\Gamma(k_2)}(\lambda(\varphi))^{k_2}$, where $\Gamma(\cdot)$ is the common Gamma function, μ is the average synthesis rate of X (given in molar concentration per second), and two parameters k_1 and k_2 are positive numbers, which may be positive real numbers in theory and can be used to model the numbers of reaction steps involved in the considered processes [45]. Function $\lambda(\varphi) = \delta\varphi^n$ may be regarded as the reaction propensity function in the Markovian reaction case, where δ represents the mean degradation rate of X. With these settings, we can show that Eq. (4) reduces to

$$W_{1}(\varphi) = \frac{(\mu)^{k_{1}} \sum_{i=0}^{k_{2}-1} {\binom{k_{1}+i-1}{i}} \frac{(\delta\varphi^{n})^{i}}{(\mu+\delta\varphi^{n})^{k_{1}+i}}}{\tau_{g} + \sum_{i=0}^{k_{1}-1} \sum_{j=0}^{k_{2}-1} {\binom{i+j}{i}} \frac{(\mu)^{i}(\delta\varphi^{n})^{j}}{(\mu+\delta\varphi^{n})^{i+j+1}}},$$

$$W_{2}(\varphi) = \frac{(\delta\varphi^{n})^{k_{2}} \sum_{i=0}^{k_{1}-1} {\binom{k_{2}+i-1}{i}} \frac{(\mu)^{i}}{(\mu+\delta\varphi^{n})^{k_{2}+i}}}{\tau_{g} + \sum_{i=0}^{k_{1}-1} \sum_{j=0}^{k_{2}-1} {\binom{i+j}{i}} \frac{(\mu)^{i}(\delta\varphi^{n})^{j}}{(\mu+\delta\varphi^{n})^{i+j+1}}},$$
(13)

due to $\int_{t}^{\infty} \alpha_{1}^{(r)}(t'; \varphi) dt' = e^{-\mu t} \sum_{i=0}^{k_{1}-1} (t\mu)^{i}/i!$ and $\int_{t}^{\infty} \alpha_{2}^{(r)}(t'; \varphi) dt' = e^{-\delta \varphi^{n} t} \sum_{i=0}^{k_{2}-1} (t\delta \varphi^{n})^{i}/i!$. The deterministic equation is given by $d\varphi/dt = mW_{1}(\varphi) - nW_{2}(\varphi)$, and the steady state $(\tilde{\varphi})$ is given by solving algebraic equation $mW_{1}(\tilde{\varphi}) - nW_{2}(\tilde{\varphi}) = 0$. Using the above notations, we know vector $\mathbf{S} = (m - n)$, matrix $\mathbf{A} = mW'_{1}(\tilde{\varphi}) - nW'_{2}(\tilde{\varphi})$, where "W'(x)" represents the derivative of function W(x) with regard to x, and matrix $\mathbf{BB}^{\mathrm{T}} = m^{2}W_{1}(\varphi) + n^{2}W_{2}(\varphi)$. From Lyapunov matrix equation $\mathbf{A\Sigma} + \Sigma \mathbf{A}^{\mathrm{T}} + \Omega \mathbf{BB}^{\mathrm{T}} = \mathbf{0}$, we can derive

$$\Sigma = -\frac{\Omega}{2} \frac{m^2 W_1(\tilde{\varphi}) + n^2 W_2(\tilde{\varphi})}{m W'_1(\tilde{\varphi}) - n W'_2(\tilde{\varphi})}.$$
 (14a)

Thus, the noise intensity for X is given by

$$\eta_X = \frac{1}{2\Omega\bar{\varphi}^2} \frac{m^2 W_1(\tilde{\varphi}) + n^2 W_2(\tilde{\varphi})}{-mW'_1(\tilde{\varphi}) + nW'_2(\tilde{\varphi})},\tag{14b}$$

	Macroscopic mean: $\Omega \bar{\varphi}$	Fano factor: Fano	Noise intensity: η_X
Case 1: $k_1 > 0,$	$ ilde{arphi} = \left\{ \left[\left(rac{m+n}{n} ight)^{1/k_1} - 1 ight] ilde{\mu} ight\}^{1/n}$	$\frac{m}{2nk_1} \frac{(m+n)^{1/k_1}}{(m+n)^{1/k_1} - n^{1/k_1}}$	$\frac{m}{2nk_1\Omega\bar{\varphi}}\frac{(m+n)^{1/k_1}}{(m+n)^{1/k_1}-n^{1/k_1}}$
$k_2 = 1$ Case 2: $k_1 = 1$, $k_2 > 0$	$\tilde{\varphi} = \left(\frac{\tilde{\mu}m^{1/k_2}}{(m+n)^{1/k_2}-m^{1/k_2}}\right)^{1/n}$	$\frac{1}{2k_2} \frac{(m+n)^{1/k_2}}{(m+n)^{1/k_2} - m^{1/k_2}}$	$\frac{1}{2k_2\Omega\tilde{\varphi}}\frac{(m+n)^{1/k_2}}{(m+n)^{1/k_2}-m^{1/k_2}}$
Case 3: $k_1 > 0,$ $k_2 = 2$	$\tilde{\varphi}$ is determined by $(1 + \frac{\tilde{\varphi}^n}{\tilde{\mu}})^{k_1} = \frac{m+n}{n}(1 + \frac{k_1\tilde{\varphi}^n}{\tilde{\mu} + \tilde{\varphi}^n})$	$\frac{m(\tilde{\mu}+\tilde{\varphi}^{n})}{2nk_{1}(k_{1}+1)}(\frac{\tilde{\mu}}{\tilde{\varphi}^{n}}+k_{1}+1)$	$\frac{\frac{m}{2n\Omega\tilde{\varphi}}\frac{\tilde{\mu}+\tilde{\varphi}^{n}}{k_{1}(k_{1}+1)}}{\times(\frac{\tilde{\mu}}{\tilde{\varphi}^{n}}+k_{1}+1)}$
Case 4: $k_1 = 2, k_2 > 0$	$\tilde{\varphi}$ is determined by $(1 + \frac{\tilde{\varphi}^n}{\tilde{\mu}})^{k_2} = \frac{m+n}{m}(1 + \frac{k_2\tilde{\varphi}^n}{\tilde{\mu}+\tilde{\varphi}^n})$	$rac{ar{\mu}+ar{arphi}^n}{2k_2(k_2+1)}(rac{ar{\mu}}{ar{arphi}^n}+k_2+1)$	$\frac{\frac{1}{2\Omega\tilde{\varphi}}\frac{\tilde{\mu}+\tilde{\varphi}^{n}}{k_{2}(k_{2}+1)}}{\times(\frac{\tilde{\mu}}{\tilde{\varphi}^{n}}+k_{2}+1)}$

TABLE I. Analytical results for the generalized birth-death process, where Ω is the system volume, $\bar{\varphi}$ is the microscopic mean of reactive species *X*, and $\tilde{\mu}$ is the transcription rate normalized by degradation rate δ , i.e., $\tilde{\mu} = \mu/\delta$.

and the Fano factor by

Fano =
$$\frac{1}{2\bar{\varphi}} \frac{m^2 W_1(\tilde{\varphi}) + n^2 W_2(\tilde{\varphi})}{-m W'_1(\tilde{\varphi}) + n W'_2(\tilde{\varphi})}.$$
 (14c)

In order to show the explicit effect of molecular memory, however, we consider four special cases: Case 1: $k_1 > 0$ and $k_2 = 1$; case 2: $k_1 = 1$ and $k_2 > 0$; case 3: $k_1 > 0$ and $k_2 = 2$; and case 4: $k_1 = 2$ and $k_2 > 0$. Main results are listed in Table I. Note that if one of memory indices k_1 and k_2 is more than 1, the corresponding case leads to non-Markovian dynamics. If both k_1 and k_2 are equal to 1, this case (i.e., the Markovian case) has been previously studied [6,39].

From Table I, we observe: (1) the macroscopic mean, the Fano factor, and the noise intensity for reactive species X are all independent of the mean of global stochastic reaction delay; (2) if $k_1 = k_2 = 1$, which corresponds to the Markovian case, our results reproduce previous results; (3) similar to the common birth-death process, the Fano factor is independent of the ratio of the mean transcription rate over the mean degradation rate for cases 1 and 2, but depends on stoichiometries m and n.

Next, we perform numerical calculations. Numerical results are demonstrated in Fig. 2, where Fig. 2(a) shows the dependence of the X Fano factor on memory index k_1 for two different yet fixed values of memory index k_2 whereas Fig. 2(b) shows the dependence of the Fano factor on k_2 for two different values of k_1 . From Fig. 2(a), we observe that the Fano factor is a monotonically decreasing function of memory index k_1 , but the larger the k_2 is, the more remarkable is the reduction degree. Similarly, the Fano factor is also a monotonically decreasing function of memory index k_2 , but the larger the k_1 is, the more remarkable is the reduction degree. Since the Fano factor can be a sign of whether or not a signal is amplified (specifically, a signal is amplified if its Fano factor is more than 1 and reduced if the Fano factor is less than 1), Figs. 2(a) and 2(b) imply that a larger number of reaction steps in the birth (or death) process reduces a signal

more, in agreement with intuition. From Fig. 2(b), we observe a particularly interesting fact: for a fixed value of k_1 , the Fano factor is always larger than 1, whatever the value of $k_2 \ge 1$.

Figures 2(c) or 2(d) demonstrate the dependence of the X noise (quantified by η_X) on memory index k_1 or k_2 . From Fig. 2(c), we observe that η_X is a monotonically increasing function of k_1 , and moreover, the greater the k_2 is, the more remarkable is the increasing degree. This implies that a multistep birth process always enlarges the X noise but this amplification will be decreased with increasing the number



FIG. 2. Effects of the number of reaction steps on the gene expression noise in the generalized birth-death process: (a) dependence of the X Fano factor (*Fano*) on k_1 for fixed $k_2 = 1$ (line with stars) and $k_2 = 2$ (line with circles), respectively; (b) dependence of the X Fano factor on k_2 for fixed $k_1 = 1$ (line with stars) and $k_1 = 2$ (line with circles), respectively. The other parameter values are set as $\mu = 10^{-6}$ (mol/s), $\delta = 10^{6}$ (s⁻¹), $\tau_g = 3$ (s), m = 4, n = 2, and $\Omega = 10^{-15}$ (liter).

of reaction steps in the death process. Similarly, η_X is a monotonically decreasing function of memory index k_1 , but the decreasing degree does not apparently depend on the value of k_1 [Fig. 2(d)]. This hints that the X noise always reduces with increasing the number of reaction steps in the death stage but the number of reaction steps in the birth process does not impact this reduction remarkably.

In summary, molecular memory is an un-neglectable factor for affecting the Fano factor and the noise in the birth-death process: It can not only amplify a signal in its propagation process but can also reduce its noise. This property would be utilized by cells or living organisms survived in complex environments.

B. Zero-order sensitivity and fluctuation size: Analysis of a generalized modification-demodification system

Two-component switching systems can find their prototypes in many scientific fields, e.g., genetic toggle-switch networks in the biological field. Such a kind of systems can model the cross repression between the determinants of different cellular states, which can result in a definite choice between two outcomes [46–48]. Conventional models of genetic toggle switch consider exponential waiting-time distributions. However, the expression of a gene in general involves a multistep process. Indeed, transcriptional repressor monomer (X_1 or X_2) binds first to dimers and then to specific DNA sequences near the promoter, repressing the production of transcriptional repressor monomer (X_2 or X_1). This multistep process can lead to nonexponential waiting times, creating a memory between individual reaction events.

Another similar example is a generalized modificationdemodification system. As pointed out by Goldbeter and Koshland [49], the extent of modification in certain enzymatically catalyzed modification-demodification reactions can be ultrasensitive to the amount of either the modifying or the demodifying enzyme. It was also shown that such a system can display large fluctuations, which will tend to attenuate the macroscopically predicted ultrasensitivity [50]. We exemplify this by treating the case with a protein that is converted from its unmodified state X_1 to its modified state X_2 by one enzyme and back to X_1 from X_2 by another, where the converting processes would be multistep. With a constant total concentration φ_0 of modified and unmodified target concentrations, the state of the system is completely described by the concentration φ_1 of X_1 or by the concentration φ_2 of X_2 due to the conservation condition: $\varphi_0 = \varphi_1 + \varphi_2$.

The above examples or other similar systems can be unitedly treated as the following generalized model of toggle switch with two reactions:

$$X_1 \xrightarrow{\alpha_1^{(r)}(t;\boldsymbol{\varphi})} X_2, \quad X_2 \xrightarrow{\alpha_2^{(r)}(t;\boldsymbol{\varphi})} X_1, \quad \boldsymbol{\varphi} = (\varphi_1, \varphi_2)^{\mathrm{T}}.$$
 (15)

We assume that $\alpha_1^{(r)}(t; \boldsymbol{\varphi}) = \frac{t^{k_1-1}e^{-\mu\varphi_1 t}}{\Gamma(k_1)} (\mu\varphi_1)^{k_1}$ and $\alpha_2^{(r)}(t; \boldsymbol{\varphi}) = \frac{t^{k_2-1}e^{-\delta\varphi_2 t}}{\Gamma(k_2)} (\delta\varphi_2)^{k_2}$ with k_1 and k_2 being positive numbers (remark: k_1 and k_2 may be two positive real numbers in theory). According to Eq. (4), it is not difficult to show that

two effective transition rates are, respectively, given by

$$W_{1}(\boldsymbol{\varphi}) = \frac{(\mu\varphi_{1})^{k_{1}}\sum_{i=0}^{k_{2}-1} {\binom{k_{1}+i-1}{i}} \frac{(\delta\varphi_{2})^{i}}{(\mu\varphi_{1}+\delta\varphi_{2})^{k_{1}+i}}}{\tau_{g} + \sum_{i=0}^{k_{1}-1}\sum_{j=0}^{k_{2}-1} {\binom{i+j}{i}} \frac{(\mu\varphi_{1})^{i}(\delta\varphi_{2})^{j}}{(\mu\varphi_{1}+\delta\varphi_{2})^{i+j+1}}},$$

$$W_{2}(\boldsymbol{\varphi}) = \frac{(\delta\varphi_{2})^{k_{2}}\sum_{i=0}^{k_{1}-1} {\binom{k_{2}+i-1}{i}} \frac{(k_{2}+i-1)}{(\mu\varphi_{1}+\delta\varphi_{2})^{k_{2}+i}}}{\tau_{g} + \sum_{i=0}^{k_{1}-1}\sum_{j=0}^{k_{2}-1} {\binom{i+j}{i}} \frac{(\mu\varphi_{1})^{i}(\delta\varphi_{2})^{j}}{(\mu\varphi_{1}+\delta\varphi_{2})^{i+j+1}}}.$$
(16)

The deterministic equation takes the form $d\varphi_1/dt = -W_1(\varphi) + W_2(\varphi)$, where $\varphi_1 = \varphi_0 - \varphi_2$ with φ_0 being a positive constant. From $\mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}} + \Omega\mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{0}$, we can obtain

$$\begin{split} \Sigma_{11} &= \frac{\Omega}{2} \frac{W_1(\tilde{\boldsymbol{\varphi}}) + W_2(\tilde{\boldsymbol{\varphi}})}{\partial_{\varphi_1} W_1(\tilde{\boldsymbol{\varphi}}) - \partial_{\varphi_1} W_2(\tilde{\boldsymbol{\varphi}})}, \\ \Sigma_{22} &= \frac{\Omega}{2} \frac{W_1(\tilde{\boldsymbol{\varphi}}) + W_2(\tilde{\boldsymbol{\varphi}})}{\partial_{\varphi_2} W_2(\tilde{\boldsymbol{\varphi}}) - \partial_{\varphi_2} W_1(\tilde{\boldsymbol{\varphi}})}, \end{split}$$

where $\tilde{\varphi}$ is the macroscopic steady state of the system. The noise intensity for X_2 is then given by

$$\eta_2 = \frac{1}{2\Omega\tilde{\varphi}_2^2} \frac{W_1(\tilde{\boldsymbol{\varphi}}) + W_2(\tilde{\boldsymbol{\varphi}})}{-\partial_{\varphi_1}W_2(\tilde{\boldsymbol{\varphi}}) + \partial_{\varphi_1}W_1(\tilde{\boldsymbol{\varphi}})}.$$
(17)

For clarity, we consider the following two special cases: Case 1: $k_1 > 0$ and $k_2 = 1$.

If the steady state is denoted by $\tilde{\boldsymbol{\varphi}} = (\tilde{\varphi}_1, \tilde{\varphi}_2)^{\mathrm{T}}$, then we have (seeing Appendix D for derivation)

$$\tilde{\varphi}_1 = \frac{\varphi_0}{(2^{1/k_1} - 1)\tilde{\mu} + 1}, \quad \tilde{\varphi}_2 = \frac{(2^{1/k_1} - 1)\tilde{\mu}\varphi_0}{(2^{1/k_1} - 1)\tilde{\mu} + 1}, \quad (18)$$

where $\tilde{\mu} = \mu/\delta$. We can show that the noise intensity is given by

$$\eta_{2} = \frac{1}{2\Omega\tilde{\varphi}_{2}^{2}} \frac{W_{1}(\tilde{\varphi}) + W_{2}(\tilde{\varphi})}{\partial_{\varphi_{1}}W_{1}(\tilde{\varphi}) - \partial_{\varphi_{1}}W_{2}(\tilde{\varphi})} \\ = \frac{1}{2k_{1}\Omega\tilde{\mu}\varphi_{0}} \frac{1}{(2^{1/k_{1}} - 1)(1 - 2^{-1/k_{1}})},$$
(19)

which is apparently independent of stochastic delay.

Case 2: $k_1 = 1$ and $k_2 > 0$.

The steady state is determined by $(\tilde{\mu}\tilde{\varphi}_1)$ $[((\tilde{\mu}-1)\tilde{\varphi}_1+\varphi_0)^{k_2}-(\varphi_0-\tilde{\varphi}_1)^{k_2}]=(\tilde{\mu}\tilde{\varphi}_1)(\varphi_0-\tilde{\varphi}_1)^{k_2}$, which gives

$$\tilde{\varphi}_2 = \frac{\tilde{\mu}\varphi_0}{2^{1/k_2} - 1 + \tilde{\mu}}.$$
(20)

The X_2 noise intensity is given by (also seeing Appendix D)

$$\eta_{2} = \frac{1}{2\Omega\tilde{\varphi}_{2}^{2}} \frac{W_{1}(\tilde{\varphi}) + W_{2}(\tilde{\varphi})}{\partial_{\varphi_{1}}W_{1}(\tilde{\varphi}) - \partial_{\varphi_{1}}W_{2}(\tilde{\varphi})} \\ = \frac{1}{2k_{2}\Omega\tilde{\mu}\varphi_{0}} \frac{2^{1/k_{2}} - 1 + \tilde{\mu}}{[(\tilde{\mu} - 1)2^{-1/k_{2}} + 1]},$$
(21)

which is also independent of stochastic reaction delay.

Next, we carry out numerical simulations with results shown Fig. 3, where Figs. 3(a) and 3(b) demonstrate how the mean concentration of the X_2 reactive species, $\tilde{\varphi}_2$, depends on memory index k_1 for a fixed value of k_2 and on k_2 for a



FIG. 3. Effects of the number of reaction steps on properties of the modification-demodification system (one-component system): (a) dependence of average concentration $\tilde{\varphi}_2$ (unit: molar) on k_1 for fixed $k_2 = 1$; (b) dependence of $\tilde{\varphi}_2$ on k_2 for fixed $k_1 = 1$; (c) dependence of the X_2 Fano factor on k_1 for fixed $k_2 = 1$; (d) dependence of the X_2 Fano factor on k_2 for fixed $k_1 = 1$; (e) dependence of the X_2 noise on k_1 for fixed $k_2 = 1$; (d) dependence of the X_2 noise on k_1 for fixed $k_2 = 1$; (d) dependence of the X_2 noise on k_2 for fixed $k_1 = 1$. The other parameter values are set as $\mu = 0.1$ (mol/s), $\delta = 0.05$ (mol/s), $\tau_g = 3$ (s), (the total concentration) $\varphi_0 = 10^{-5}$ (mol), and $\Omega = 10^{-15}$ (liter). We point out that the case of $k_1 < 1$ or $k_2 < 1$ in (c)–(f) is not shown since the noise intensity or the Fano factor is very small when the memory index is less than 1.

fixed value of k_1 , respectively. We observe that when the total concentration of X_1 and X_2 is fixed, $\tilde{\varphi}_2$ first changes slowly and then drops down sharply and finally tends to a small value with increasing k_1 . The change tendency in the k_2 case is contrary to that in the k_1 case. These two diagrams imply that molecular memory can induce (zero-order) ultrasensitivity of reactive species by remarkably changing their concentrations.

Figures 3(c) and 3(d) show how the X_2 Fano factor depends on memory indexes k_1 and k_2 , respectively. We observe that this factor is a monotonically increasing (decreasing) function of k_1 (k_2), implying that k_1 amplifies the X_2 signal, whereas k_2 reduces the X_2 signal. Similar change tendency holds for the X_2 noise.

Numerical results shown in Fig. 3 imply that for the modification-demodification system with a conservative total concentration, molecular memory can induce ultrasensitivity of a reactive species, and enlarge or reduce the noise in the species, depending on the source of molecular memory. This property would also be utilized by cells or living organisms survived in changing environments.

IV. DISCUSSION

We have extended the LNAs in Markovian reaction cases to the gLNAs (or the gFDTs) in non-Markovian reaction cases (i.e., the above gFDT theorem). This extension can be conveniently used in the fast evaluation of stationary fluctuations in the concentrations of reactive species in general reaction networks with arbitrary reaction waiting-time distributions. In particular, it can provide insights into how molecular memory fine-tunes the noise in reactive species of interest in the complex non-Markovian reaction networks, as shown in the above. We believe that like the common LNA, the gLNA or the gFDT can also have broad applications although we analyzed only two examples in this paper.

Tracing sources of noise in the output of a reaction network is an interesting topic. This involves the question of their different contributions, e.g., how does the noise from each reaction in the network, including the noise generated due to the distributed reaction waiting times, differently contribute to the total noise or the output noise? From the gFDT derived above, we can also give the noise composition principle mainly because the effect of molecular memory has been explicitly incorporated in the effective transition rates. For clarity, let us consider the above birth-death process in the case of m = n = 1 and $k_1 > 0$, $k_2 = 1$. In this case, we can show that the noise intensity is given by

$$\eta_X = \frac{\delta}{\Omega\mu} \frac{1/(2k_1)}{(2^{1/k_1} - 1)(1 - 2^{-1/k_1})},\tag{22}$$

where $\delta/(\Omega\mu)$ represents the noise intensity in the Markovian reaction case, and factor $(1/(2k_1))/[(2^{1/k_1} - 1)(1 - 2^{-1/k_1})]$ appears due to the introduction of nonexponential waiting-time distribution for the synthetic reaction. In other words, molecular memory impacts the output noise in a nonlinear manner. This qualitative result seems universal.

Recall that Franzke and Kosko's Markov chain noise benefit theorem (MCNBT) showed that for a Markovian reaction system, intermediate noise levels-similarly to a stochastic resonance effect-can speed up the convergence of the system to a steady state [51,52]. However, for non-Markovian reaction systems, the MCNBT would not hold. For this, let us still consider the previous non-Markovian birth-death process of m = n = 1. We can show that the effect of molecular memory is equivalent to the introduction of feedback. In fact, if $\alpha_1(t, x) = [(\mu t)^{k_1 - 1} / \Gamma(k_1)] e^{-\mu t}$, where μ is the average production rate and k_1 a positive integer (modeling the step number of a multistep process and representing the memory strength), and $\alpha_2(t;x) = x\delta e^{-x\delta}$, where δ is the average decay rate and x the molecular number of reactive species X, the effect of molecular memory is equivalent to the effect of a negative feedback. In fact, according to Eq. (4) above, we can have $W_1(x) = x\delta\mu^{k_1}/[(\mu + x\delta)^{k_1} - \mu^{k_1}]$ and $W_2(x) = x\delta$. To see clearly the effect of molecular memory, we set $k_1 = 2$. Then, $W_1(x) = \mu/(2\mu + x\delta)$, indicating that the equivalent Markovian reaction system has a negative feedback. Completely similarly, if $\alpha_1(t, x) = \mu e^{-\mu t}$ and $\alpha_2(t, x) = [(x \delta t)^{k_2 - 1} / \Gamma(k_2)] e^{-x \delta t}$, where k_2 is a positive integer, we can show that the effect of molecular memory is equivalent to that of a positive feedback. On the other hand, it is well known that negative feedback reduces the noise whereas positive feedback amplifies the noise. Thus, properties of the equivalent reaction network depend on the settings of waiting-time distributions. We then conclude that for non-Markovian reaction systems, the MCNBT depends on waiting-time distributions.

Finally, we point out that the derived-above gFDT for a general reaction network is based on the assumption that inter-reaction waiting times are reset. From the viewpoint of CTRW, this resetting corresponds to the so-called active

CTRW [53]. However, there is another kind of CTRW where waiting times cannot be reset, e.g., those in the sense of queuing theory [54]. Which kind of waiting time is used, the more reasonable modeling of biochemical reaction networks is unclear. And, the relationship between the two classes of waiting times has not been established yet. In addition, Jia and Kulkarni studied a queuing model of stochastic gene expression and derived an estimate for the noise in steadystate protein distributions [55], and Pedraza and Paulson [7] analyzed another queuing model of stochastic gene expression and presented another estimate for the expression noise, which formulates the variance in protein abundance in terms of the randomness of the individual gene expression events. In spite of these, the question of how the gFDT is established for queuing models of biochemical reaction networks is an issue worth exploring.

This work was supported by Grants No. 11931019 and No. 11775314 from Natural Science Foundation of People's Republic of China, by Grant No. 202007030004 from Key-Area Research and Development Program of Guangzhou, People's Republic of China, and by the Hubei Province Education Department under Grant No. B2020074.

APPENDIX A: DERIVATION OF EFFECTIVE TRANSITION RATES

Denote by $\bar{\kappa}_i(s, \boldsymbol{x}; \Omega)$ the Laplace transform of $\kappa_i(t, \boldsymbol{x}; \Omega)$ in the main text. According to the property of Laplace transform, we have $\bar{\kappa}_i(s, \boldsymbol{x}; \Omega) = \bar{\kappa}_i^{(r)}(s, \boldsymbol{x}; \Omega) \bar{\kappa}_i^{(d)}(s, \boldsymbol{x}; \Omega)$. According to the definition of $\bar{M}_j(\boldsymbol{x}; \Omega)$ in the main text,

we have

$$\begin{split} \bar{M}_j(s, \boldsymbol{x}; \Omega) &= \frac{s\bar{\kappa}_j(s, \boldsymbol{x}; \Omega)}{1 - \sum_{j=1}^L \bar{\kappa}_j(s, \boldsymbol{x}; \Omega)} = \frac{s\bar{\kappa}_j^{(d)}(s, \boldsymbol{x}; \Omega)\bar{\kappa}_j^{(r)}(s, \boldsymbol{x}; \Omega)}{1 - \sum_{j=1}^L \bar{\kappa}_j^{(d)}(s, \boldsymbol{x}; \Omega)\bar{\kappa}_j^{(r)}(s, \boldsymbol{x}; \Omega)} \\ &= \frac{s\int_0^\infty e^{-st}\kappa_j^{(d)}(t, \boldsymbol{x}; \Omega)dt\int_0^\infty e^{-st}\kappa_j^{(r)}(t, \boldsymbol{x}; \Omega)dt}{1 - \sum_{j=1}^L \int_0^\infty e^{-st}\kappa_j^{(d)}(t, \boldsymbol{x}; \Omega)dt\int_0^\infty e^{-st}\kappa_j^{(r)}(t, \boldsymbol{x}; \Omega)dt}. \end{split}$$

1

Note that

$$\frac{d}{ds}\left\{1-\sum_{j=1}^{L}\int_{0}^{\infty}e^{-st}\kappa_{j}^{(d)}(t,\boldsymbol{x};\Omega)dt\int_{0}^{\infty}e^{-st}\kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega)dt\right\}$$
$$=s\sum_{j=1}^{L}\int_{0}^{\infty}te^{-st}\kappa_{j}^{(d)}(t,\boldsymbol{x};\Omega)dt\int_{0}^{\infty}e^{-st}\kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega)dt$$
$$+s\sum_{j=1}^{L}\int_{0}^{\infty}e^{-st}\kappa_{j}^{(d)}(t,\boldsymbol{x};\Omega)dt\int_{0}^{\infty}te^{-st}\kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega)dt.$$

By L'Hospital rule, we have

$$\begin{split} W_{i}(\boldsymbol{x};\Omega) &= \lim_{\tau \to 0} \bar{M}_{j}(\tau,\boldsymbol{x};\Omega) \\ &= \frac{\int_{0}^{\infty} \kappa_{j}^{(d)}(t,\boldsymbol{x};\Omega) dt \int_{0}^{\infty} \kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega) dt}{\sum_{j=1}^{L} \left\langle \tau_{j}^{(d)} \right\rangle \int_{0}^{\infty} \kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega) dt + \sum_{j=1}^{L} \int_{0}^{\infty} \kappa_{j}^{(d)}(t,\boldsymbol{x};\Omega) dt \int_{0}^{\infty} t \kappa_{j}^{(r)}(t,\boldsymbol{x};\Omega) dt}, \end{split}$$

where $\langle \tau_i^{(d)} \rangle = \int_0^\infty t \kappa_i^{(d)}(t, \mathbf{x}; \Omega) dt$. Note that $\int_0^\infty \kappa_i^{(d)}(t, \mathbf{x}; \Omega) dt \equiv 1$ and

$$\begin{split} \sum_{j=1}^{L} \int_{0}^{\infty} \kappa_{j}^{(d)}(t, \boldsymbol{x}; \Omega) dt \int_{0}^{\infty} t \kappa_{j}^{(r)}(t, \boldsymbol{x}; \Omega) dt \\ &= \sum_{j=1}^{L} \int_{0}^{\infty} t \alpha_{j}^{(r)}(t, \boldsymbol{x}; \Omega) \prod_{i \neq j} \int_{t}^{+\infty} \alpha_{i}^{(r)}(t', \boldsymbol{x}; \Omega) dt' dt \\ &= -\int_{0}^{+\infty} t \left[\frac{d}{dt} \prod_{j=1}^{L} \int_{t}^{+\infty} \alpha_{i}^{(r)}(t', \boldsymbol{x}; \Omega) dt' \right] dt = \int_{0}^{+\infty} \prod_{j=1}^{L} \int_{t}^{+\infty} \alpha_{i}^{(r)}(t', \boldsymbol{x}; \Omega) dt' dt. \end{split}$$

Thus, we obtain

$$W_i(\boldsymbol{x};\Omega) = \frac{\int_0^\infty \kappa_j^{(r)}(t,\boldsymbol{x};\Omega)dt}{\sum_{j=1}^L \langle \tau_j^{(d)} \rangle \int_0^\infty \kappa_j^{(r)}(t,\boldsymbol{x};\Omega)dt + \int_0^{+\infty} \prod_{j=1}^L \int_t^{+\infty} \alpha_i^{(r)}(t',\boldsymbol{x};\Omega)dt'dt}.$$
(A1)

According to the definition of function $W_j(\varphi)$, i.e., $W_j(\varphi) = \lim_{\Omega \to \infty} W_j(\mathbf{x}; \Omega)$, and note the fact that $\sum_{j=1}^{L} \int_0^{\infty} \kappa_j^{(r)}(t, \mathbf{x}; \Omega) dt = 1$, we finally obtain Eq. (4) in the main text.

APPENDIX B: PROOF OF THE GENERALIZED FLUCTUATION THEOREM

In general, we can write $X = \Omega x + \Omega^{1/2} \xi$, where ξ is a vector of random variables. Under this transform, the original stationary PDF, P(X) becomes a new PDF, denoted by $\Pi(\xi)$. Then, we perform the following Ω expansions: (I) expand function $W_i(x; \Omega)$ into

$$W_j(\boldsymbol{x}; \Omega) = W_j(\boldsymbol{\varphi} + \Omega^{-1/2}\boldsymbol{\xi})$$

= $W_j(\boldsymbol{\varphi}) + \Omega^{-1/2} \sum_{i=1}^N \boldsymbol{\xi}_i \frac{\partial W_j(\boldsymbol{\varphi})}{\partial x_i} + O(\Omega^{-1})$ (B1)

(II) For the step operator in the stationary gCME, i.e., Eq. (7) in the main text, we in general have the approximation

$$\mathbb{E}^{\beta}f(x) = f(x + \Omega^{-1/2}(\xi + \Omega^{-1/2}\beta))$$
$$= \left[1 + \Omega^{-1/2}\beta\frac{\partial}{\partial\xi} + \frac{\Omega^{-1/2}\beta^2}{2}\frac{\partial^2}{\partial\xi\partial\xi} + O(\Omega^{-3/2})\right]f(x).$$
(B2)

where β is a real number. For jumps in multiple dimensions, we have

$$\sum_{i=1}^{N} \mathbb{E}^{-S_{ij}} - \mathbb{I} = -\Omega^{-1/2} \sum_{i=1}^{N} S_{ij} \frac{\partial}{\partial \xi_i} + \frac{1}{2} \Omega^{-1} \sum_{i,k=1}^{N} S_{ij} S_{kj} \frac{\partial^2}{\partial \xi_i \partial \xi_k} + O(\Omega^{-3/2}).$$
(B3)

Inserting Eqs. (B1)–(B3) into Eq. (7) in the main text yields

$$0 = -\Omega^{-1/2} \sum_{j=1}^{L} \sum_{i=1}^{N} S_{ij} W_j(\boldsymbol{\varphi}) \frac{\partial \Pi(\boldsymbol{\xi})}{\partial \xi_i} + \Omega^{-1} \sum_{j=1}^{L} \left[\frac{1}{2} \sum_{i,k=1}^{N} S_{ij} S_{kj} W_j(\boldsymbol{\varphi}) \frac{\partial^2 \Pi(\boldsymbol{\xi})}{\partial \xi_i \partial \xi_k} - \sum_{i=1}^{N} \sum_{\ell=1}^{N} S_{ij} \frac{\partial W_j(\boldsymbol{\varphi})}{\partial x_\ell} \frac{\partial (\xi_\ell \Pi(\boldsymbol{\xi}))}{\partial \xi_i} \right] + O(\Omega^{-3/2}).$$

Identifying the term of order $\Omega^{-1/2}$ gives

$$\sum_{j=1}^{L} \sum_{i=1}^{N} S_{ij} W_j(\boldsymbol{\varphi}) \frac{\partial \Pi(\boldsymbol{\xi})}{\partial \xi_i} = 0, \qquad (B4)$$

which naturally holds due to Eq. (8) in the main text. Identifying the term of order Ω^{-1} yields

$$\frac{1}{2} \sum_{j=1}^{L} \sum_{i,k=1}^{N} S_{ij} S_{kj} W_j(\boldsymbol{\varphi}) \frac{\partial^2 \Pi(\boldsymbol{\xi})}{\partial \xi_i \partial \xi_k} \\ - \sum_{j=1}^{L} \sum_{i=1}^{N} \sum_{\ell=1}^{N} S_{ij} \frac{\partial W_j(\boldsymbol{\varphi})}{\partial x_\ell} \frac{\partial (\xi_\ell \Pi(\boldsymbol{\xi}))}{\partial \xi_i} = 0,$$

which can be further rewritten as

$$-\sum_{i,j=1}^{N} A_{ij} \frac{\partial [\xi_j \Pi(z)]}{\partial \xi_i} + \frac{1}{2} \sum_{i,j=1}^{N} [\mathbf{B}\mathbf{B}^{\mathsf{T}}]_{ij} \frac{\partial^2 \Pi(z)}{\partial \xi_i \partial \xi_j} = 0, \quad (\mathsf{B5a})$$

where

$$A_{ij} = \sum_{k=1}^{L} S_{ik} \frac{\partial W_k(\boldsymbol{\varphi})}{\partial \varphi_j} = \frac{\partial [\mathbf{S}_i \boldsymbol{W}(\boldsymbol{\varphi})]}{\partial \varphi_j}, \quad (B5b)$$
$$[\mathbf{B}\mathbf{B}^{\mathrm{T}}]_{ij} = \sum_{k=1}^{L} S_{ik} S_{jk} W_k(\boldsymbol{\varphi})$$
$$= [\mathbf{S}^{\mathrm{T}} \mathrm{diag}(W_1(\boldsymbol{\varphi}), \dots, W_L(\boldsymbol{\varphi})) \mathbf{S}]_{ij}. \quad (B5c)$$

Note that Eq. (B5a) is a stationary linear Fokker-Plank equation [56] with coefficient matrices **A** and **BB**^T that will be assumed to be evaluated in the macroscopic stationary state $\bar{\varphi}$. The stationary solution of the linear Fokker-Planck equation, i.e., Eq. (B5a), is a multidimensional normal distribution of the following form [54,57]:

$$\Pi(\boldsymbol{\xi}) = \frac{1}{\sqrt{(2\pi)^N \det\left(\boldsymbol{\Sigma}\right)}} \exp\left(-\frac{1}{2}\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\right), \qquad (B6)$$

where det(Σ) represents the determinant of matrix $\Sigma = \xi \xi^{T}$. This distribution has the zero average vector and the covariance matrix $\mathbf{\Sigma} = \boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}$ that follows a Lyapunov equation of the form $\mathbf{A}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}} = 0$. Note that the correlation matrix of the stationary process is given by $\boldsymbol{\xi}(t)\boldsymbol{\xi}^{\mathrm{T}}(s) =$ $\Sigma e^{|t-s|A}$. Using the relation of $X = \Omega \varphi + \Omega^{1/2} \xi$, we know that the covariance matrix $\mathbf{C} = \langle (\mathbf{X} - \langle \mathbf{X} \rangle) (\mathbf{X} - \langle \mathbf{X} \rangle)^{\mathrm{T}} \rangle =$ $\langle \Delta X \cdot \Delta X^{\mathrm{T}} \rangle$ is given by $\mathbf{C} = \Omega \boldsymbol{\Sigma}$. Thus, the algebraic equation $\mathbf{A}\Sigma + \Sigma \mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{B}^{\mathrm{T}} = 0$ can be rewritten as Eq. (10) in the main text. Matrices A and BB^T are evaluated in the macroscopic stationary state and hence constant matrices, so we can obtain the expressions for the variances of individual reactive species (i.e., for the diagonal elements of matrix C). If the noise intensity is defined as the ratio of the variance over the squared mean, we can further obtain the formal expression of the noise intensity for each individual reactive species, i.e., Eq. (11) in the main text. The above generalized fluctuation theorem provides a fast evaluation for stochastic fluctuations in individual reactive species in a reaction network. Here, we emphasize that the above generalized fluctuation theorem is derived in the non-Markovian case and although similar in the form, it is different from the traditional linear noise approximation since functions $W_k(\varphi)$ ($1 \le k \le N$) in the former have incorporated the effects of molecular memory.

APPENDIX C: DERIVATION OF THE NOISE IN THE BIRTH-DEATH PROCESS

In the following, we assume that all stochastic reaction delays $\langle \tau_j^{(d)} \rangle$ are equal, and denote by τ_g the mean of the common stochastic reaction delay.

Case 1: $k_1 > 0$ and $k_2 = 1$.

In this case, we can show that two effective transition rates are given by

$$W_1(\varphi) = \frac{\mu^{k_1} \delta \varphi^n}{(\mu + \delta \varphi^n)^{k_1} - \mu^{k_1} + \tau_g \delta \varphi^n (\mu + \delta \varphi^n)^{k_1}} \quad \text{and} \quad W_2(\varphi) = \frac{(\delta \varphi^n) [(\mu + \delta \varphi^n)^{k_1} - \mu^{k_1}]}{(\mu + \delta \varphi^n)^{k_1} - \mu^{k_1} + \tau_g \delta \varphi^n (\mu + \delta \varphi^n)^{k_1}}$$

The steady state $\tilde{\varphi}$, determined by solving algebraic equation $mW_1(\tilde{\varphi}) = nW_2(\tilde{\varphi})$, is given by

$$(\mu + \delta \tilde{\varphi}^n)^{k_1} = \frac{m+n}{n} \mu^{k_1}, \quad \text{or} \quad \tilde{\varphi} = \left\{ \left[\left(\frac{m+n}{n}\right)^{1/k_1} - 1 \right] \tilde{\mu} \right\}^{1/n}, \tag{C1}$$

indicating that the mean global delay does not impact the steady state, where $\tilde{\mu} = \mu/\delta$. Note that

 $0 \sim n$

$$\frac{W'_1(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{n}{m + \tau_g(m+n)\delta\tilde{\varphi}^n} - \frac{n(m+n)\delta\tilde{\varphi}^n[k_1 + \tau_g(k_1\delta\tilde{\varphi}^n + \mu + \delta\tilde{\varphi}^n)]}{(\mu + \delta\tilde{\varphi}^n)[m + \tau_g(m+n)\delta\tilde{\varphi}^n]^2}$$

and

$$\frac{W_2'(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{m+k_1(m+n)\left(\frac{\delta\varphi^n}{\mu+\delta\tilde{\varphi}^n}\right)}{m+\tau_g(m+n)\delta\tilde{\varphi}^n} - \frac{m(m+n)(\delta\tilde{\varphi}^n)[k_1+\tau_g(k_1\delta\tilde{\varphi}^n+\mu+\delta\tilde{\varphi}^n)]}{(\mu+\delta\tilde{\varphi}^n)[m+\tau_g(m+n)\delta\tilde{\varphi}^n]^2}.$$

Thus,

$$-\frac{mW'_1(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} + \frac{nW'_2(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{k_1n(m+n)\left(\frac{\delta\varphi^n}{\mu+\delta\tilde{\varphi}^n}\right)}{m+\tau_g(m+n)\delta\tilde{\varphi}^n}.$$
(C2)

Also note that $W_1(\tilde{\varphi}) = \frac{n\delta\tilde{\varphi}^n}{m+\tau_g(m+n)\delta\tilde{\varphi}^n}$ and $W_2(\tilde{\varphi}) = \frac{m(\delta\tilde{\varphi}^n)}{m+\tau_g(m+n)\delta\tilde{\varphi}^n}$. Therefore,

$$m^2 W_1(\tilde{\varphi}) + n^2 W_2(\tilde{\varphi}) = \frac{mn(m+n)\delta\tilde{\varphi}^n}{m + \tau_g(m+n)\delta\tilde{\varphi}^n}.$$
(C3)

According to Eq. (14a) in the main text, we thus obtain the analytical expression for the X noise:

$$\eta_X = \frac{m\tilde{\mu}}{2k_1\Omega n\tilde{\varphi}^{n+1}} \left(\frac{m+n}{n}\right)^{1/k_1} = \frac{1}{2k_1\Omega\tilde{\varphi}} \frac{m}{n} \frac{(m+n)^{1/k_1}}{(m+n)^{1/k_1} - n^{1/k_1}},\tag{C4}$$

which is independent of stochastic delay.

Case 2: $k_1 = 1$ and $k_2 > 0$.

In this case, two effective transition rates are $W_1(\varphi) = \frac{\mu[(\mu+\delta\varphi^n)^{k_2}-(\delta\varphi^n)^{k_2}]}{(\mu+\delta\varphi^n)^{k_2}-(\delta\varphi^n)^{k_2}+\tau_s\mu(\mu+\delta\varphi^n)^{k_2}}$ and $W_2(\varphi) = \frac{\mu(\delta\varphi^n)^{k_2}}{(\mu+\delta\varphi^n)^{k_2}-(\delta\varphi^n)^{k_2}+\tau_s\mu(\mu+\delta\varphi^n)^{k_2}}$. The steady state is determined by the equation $(\mu+\delta\tilde{\varphi}^n)^{k_2} = \frac{m+n}{m}(\delta\tilde{\varphi}^n)^{k_2}$ or $\tilde{\varphi} = \left[\frac{\tilde{\mu}m^{1/k_2}}{(m+n)^{1/k_2}-m^{1/k_2}}\right]^{1/n}$. Note that $W_1(\tilde{\varphi}) = \frac{n\mu}{n+\tau_s\mu(m+n)}$ and $W_2(\tilde{\varphi}) = \frac{m\mu}{n+\tau_s\mu(m+n)}$. Therefore,

$$m^2 W_1(\tilde{\varphi}) + n^2 W_2(\tilde{\varphi}) = \frac{mn(m+n)\mu}{n + \tau_g \mu(m+n)}.$$
 (C5)

Also note that

$$\frac{W'_{1}(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{1}{(\mu+\delta\tilde{\varphi}^{n})} \frac{k_{2}\mu \left[m+n-m\frac{\mu+\delta\tilde{\varphi}^{n}}{\delta\tilde{\varphi}^{n}}\right]}{n+\tau_{g}\mu(m+n)} - \frac{k_{2}\mu n\left[m+n-m\frac{\mu+\delta\tilde{\varphi}^{n}}{\delta\tilde{\varphi}^{n}}+\tau_{g}\mu(m+n)\right]}{(\mu+\delta\tilde{\varphi}^{n})[n+\tau_{g}\mu(m+n)]^{2}}$$

and

$$\frac{W_2'(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{mk_2\mu(\delta\tilde{\varphi}^n)^{-1}}{n+\tau_g\mu(m+n)} - \frac{k_2\mu m \left[m+n-m\frac{\mu+\varphi\varphi^n}{\delta\tilde{\varphi}^n}+\tau_g\mu(m+n)\right]}{(\mu+\delta\tilde{\varphi}^n)[n+\tau_g\mu(m+n)]^2}.$$

Thus,

$$-\frac{mW'_1(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} + \frac{nW'_2(\tilde{\varphi})}{n\delta\tilde{\varphi}^{n-1}} = \frac{m(m+n)k_2\mu(\delta\tilde{\varphi}^n)^{-1} - k_2\mu m\frac{m+n}{\mu+\delta\tilde{\varphi}^n}}{n+\tau_g\mu(m+n)}.$$
 (C6)

According to Eq. (14a) in the main text, the X noise is then given by

$$\eta_X = \frac{1}{2k_2 \Omega \tilde{\varphi}} \frac{(m+n)^{1/k_2}}{(m+n)^{1/k_2} - m^{1/k_2}},\tag{C7}$$

which is independent of stochastic reaction delay.

APPENDIX D: DERIVATION OF THE NOISE IN THE MODIFICATION-DEMODIFICATION SYSTEM

In the following, we assume that all stochastic reaction delays $\langle \tau_j^{(d)} \rangle$ are equal, and denote by τ_g the mean of the common stochastic reaction delay.

Case 1: $k_1 > 0$ and $k_2 = 1$. In this case, two effective transition rates are

$$W_{1}(\boldsymbol{\varphi}) = \frac{\delta(\tilde{\mu}\varphi_{1})^{k_{1}}(\varphi_{0} - \varphi_{1})}{\left\{ [(\tilde{\mu} - 1)\varphi_{1} + \varphi_{0}]^{k_{1}} - (\tilde{\mu}\varphi_{1})^{k_{1}} \right\} + \delta\tau_{g}(\varphi_{0} - \varphi_{1})[(\tilde{\mu} - 1)\varphi_{1} + \varphi_{0}]^{k_{1}}},$$
$$W_{2}(\boldsymbol{\varphi}) = \frac{\delta(\varphi_{0} - \varphi_{1})\left[[(\tilde{\mu} - 1)\varphi_{1} + \varphi_{0}]^{k_{1}} - (\tilde{\mu}\varphi_{1})^{k_{1}} \right]}{\left\{ [(\tilde{\mu} - 1)\varphi_{1} + \varphi_{0}]^{k_{1}} - (\tilde{\mu}\varphi_{1})^{k_{1}} \right\} + \delta\tau_{g}(\varphi_{0} - \varphi_{1})[(\tilde{\mu} - 1)\varphi_{1} + \varphi_{0}]^{k_{1}}}.$$

If steady state is denoted by $\tilde{\boldsymbol{\varphi}} = (\tilde{\varphi}_1, \tilde{\varphi}_2)^{\mathrm{T}}$, then we have

$$(\mu\tilde{\varphi}_1 + \delta\tilde{\varphi}_2)^{k_1} = 2(\mu\tilde{\varphi}_1)^{k_1}, \quad \text{or} \quad \tilde{\varphi}_2 = \frac{(2^{1/k_1} - 1)\tilde{\mu}\varphi_0}{(2^{1/k_1} - 1)\tilde{\mu} + 1}.$$
 (D1)

By calculation, we can obtain

$$W_{1}(\tilde{\varphi}) + W_{2}(\tilde{\varphi}) = \frac{2}{I} (\tilde{\mu}\varphi_{1})^{k_{1}} (\varphi_{0} - \varphi_{1})|_{\varphi_{1} = \tilde{\varphi}_{1}},$$
(D2)

where $I = (1/\delta) \{ [(\tilde{\mu} - 1)\varphi_1 + \varphi_0]^{k_1} - (\tilde{\mu}\varphi_1)^{k_1} \} + \tau_g(\varphi_0 - \varphi_1) [(\tilde{\mu} - 1)\varphi_1 + \varphi_0]^{k_1} \}$. Similarly,

$$-\partial_{\varphi_1} W_2(\tilde{\boldsymbol{\varphi}}) + \partial_{\varphi_1} W_1(\tilde{\boldsymbol{\varphi}}) = \frac{2k_1}{I} (2^{1/k_1} - 1) (\tilde{\mu}\varphi_1)^{k_1} [\tilde{\mu} - (\tilde{\mu} - 1)2^{-1/k_1}]|_{\varphi_1 = \tilde{\varphi}_1}.$$

Therefore, the X_2 noise intensity is given by

$$\eta_2 = \frac{1}{2\Omega\tilde{\varphi}_2^2} \frac{W_1(\tilde{\boldsymbol{\varphi}}) + W_2(\tilde{\boldsymbol{\varphi}})}{\partial_{\varphi_1} W_1(\tilde{\boldsymbol{\varphi}}) - \partial_{\varphi_1} W_2(\tilde{\boldsymbol{\varphi}})} = \frac{1}{2k_1 \Omega\tilde{\mu}\varphi_0} \frac{1}{(2^{1/k_1} - 1)(1 - 2^{-1/k_1})}.$$
(D3)

which is independent of stochastic reaction delay.

Case 2: $k_1 = 1$ and $k_2 > 0$. In this case, two effective transition rates are

$$W_{1}(\boldsymbol{\varphi}) = \frac{\delta(\tilde{\mu}\varphi_{1})[[(\tilde{\mu}-1)\varphi_{1}+\varphi_{0}]^{k_{2}}-(\varphi_{0}-\varphi_{1})^{k_{2}}]}{\{[(\tilde{\mu}-1)\varphi_{1}+\varphi_{0}]^{k_{2}}-(\varphi_{0}-\varphi_{1})^{k_{2}}\}+\delta\tau_{g}\tilde{\mu}\varphi_{1}[(\tilde{\mu}-1)\varphi_{1}+\varphi_{0}]^{k_{2}}},$$

$$W_{2}(\boldsymbol{\varphi}) = \frac{\delta(\tilde{\mu}\varphi_{1})(\varphi_{0}-\varphi_{1})^{k_{2}}}{\{[(\tilde{\mu}-1)\varphi_{1}+\varphi_{0}]^{k_{2}}-(\varphi_{0}-\varphi_{1})^{k_{2}}\}+\delta\tau_{g}\tilde{\mu}\varphi_{1}[(\tilde{\mu}-1)\varphi_{1}+\varphi_{0}]^{k_{2}}},$$

where $\tilde{\mu} = \mu/\delta$. The steady state is determined by solving the algebraic equation

$$(\tilde{\mu}\tilde{\varphi}_1)[((\tilde{\mu}-1)\tilde{\varphi}_1+\varphi_0)^{k_2}-(\varphi_0-\tilde{\varphi}_1)^{k_2}]=(\tilde{\mu}\tilde{\varphi}_1)(\varphi_0-\tilde{\varphi}_1)^{k_2}$$

which gives

$$(\tilde{\mu} - 1)\tilde{\varphi}_1 + \varphi_0 = 2^{1/k_2}(\varphi_0 - \tilde{\varphi}_1), \quad \text{or} \quad \tilde{\varphi}_2 = \frac{\tilde{\mu}\varphi_0}{2^{1/k_2} - 1 + \tilde{\mu}}$$
 (D4)

By calculation, we have

$$W_1(\tilde{\varphi}) + W_2(\tilde{\varphi}) = \frac{(2\tilde{\mu}\varphi_1)(\varphi_0 - \varphi_1)^{k_2}}{I}|_{\varphi_1 = \tilde{\varphi}_1},$$
 (D5)

where $I = \{[(\tilde{\mu} - 1)\tilde{\varphi}_1 + \varphi_0]^{k_2} - (\varphi_0 - \tilde{\varphi}_1)^{k_2}\} + \tau_g \tilde{\mu} \tilde{\varphi}_1 [(\tilde{\mu} - 1)\tilde{\varphi}_1 + \varphi_0]^{k_2}\}$. Similarly, we have

$$-\partial_{\varphi_1} W_2(\tilde{\boldsymbol{\varphi}}) + \partial_{\varphi_1} W_1(\tilde{\boldsymbol{\varphi}}) = \frac{1}{I} \{ 2k_2(\tilde{\mu}\varphi_1)(\varphi_0 - \varphi_1)^{k_2 - 1} [(\tilde{\mu} - 1)2^{-1/k_2} + 1] \}|_{\varphi_1 = \tilde{\varphi}_1}.$$
(D6)

Therefore, the X_2 noise intensity is given by

$$\eta_2 = \frac{1}{2\Omega\tilde{\varphi}_2^2} \frac{W_1(\tilde{\varphi}) + W_2(\tilde{\varphi})}{\partial_{\varphi_1} W_1(\tilde{\varphi}) - \partial_{\varphi_1} W_2(\tilde{\varphi})} = \frac{1}{2k_2 \Omega\tilde{\mu}\varphi_0} \frac{2^{1/k_2} - 1 + \tilde{\mu}}{[(\tilde{\mu} - 1)2^{-1/k_2} + 1]},$$
(D7)

which is apparently independent of stochastic reaction delay.

- J. Gunawardena, Biological systems theory, Science 328, 581 (2010).
- P. Guptasarama, Does replication-induced transcription regulate the myriad of low copy number proteins of escherichia coli? Bioessays 17, 987 (1995).
- [3] C. Gardiner, Stochastic Methods-A Handbook for the Natural and Social Sciences (Springer, New York, 2009).
- [4] N. G. van Kampen, Stochastic Processes in Physics and Chemistry (Elsevier, Amsterdam, 2007).
- [5] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 2. (John Wiley & Sons, New York, 2008).
- [6] J. J. Zhang and T. S. Zhou, Markovian approaches to modeling intracellular reaction processes with molecular memory, Proc. Natl. Acad. Sci. USA. 116, 23542 (2019).
- [7] J. M. Pedraza and P. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression, Science 319, 339 (2008).
- [8] C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, D. G. Spiller, J. J. Mullins, D. A. Rand, J. R. E. Davis, and M. R. H. White, Dynamic analysis of stochastic transcription cycles, PLoS Biol. 9, e1000607 (2011).
- [9] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, A high resolution human contact network for infectious disease transmission, Proc. Natl. Acad. Sci. USA 107, 22020 (2010).
- [10] A. Corral, Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes, Phys. Rev. Lett. 92, 108501 (2004).
- [11] P. S. Stumpf, R. C. Smith, M. Lenz, A. Schuppert, F. J. Müller, A. Babtie *et al.*, Stem cell differentiation as a non-Markov stochastic process, Cell Syst. 5, 268 (2017).
- [12] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science 332, 472 (2011).
- [13] T. Guérin, O. Bénichou, and R. Voituriez, Non-Markovian polymer reaction kinetics, Nat. Chem. 4, 568 (2012).
- [14] A. L. Barabási, The origin of bursts and heavy tails in human dynamics, Nature (London) 435, 207 (2005).
- [15] T. Aquino and M. Dentz, Chemical Continuous Time Random Walks, Phys. Rev. Lett. 119, 230601 (2017).
- [16] E. W. Montroll and H. Scher, Generalized master equations for continuous-time random walks, J. Stat. Phys. 9, 101 (1973).

- [17] U. Landman, E. W. Montroll, and M. F. Shlesinger, Random walks and generalized master equations with internal degrees of freedom, Proc. Natl. Acad. Sci. USA 74, 430 (1977).
- [18] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, Proc. Natl. Acad. Sci. USA 105, 17256 (2008).
- [19] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS Biol. 4, e309 (2006).
- [20] T. S. Zhou and J. J. Zhang, Analytical results for a multistate gene model, SIAM J. Appl. Math. 72, 789 (2012).
- [21] J. J. Zhang and T. S. Zhou, Promoter-mediated transcriptional dynamics, Biophys. J. 106, 479 (2014).
- [22] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, J. Chem. Phys. **124**, 044104 (2006).
- [23] Y. Cao and J. Liang, Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability, BMC Syst. Biol. 2, 30 (2008).
- [24] Y. Cao, H. M. Lu, and J. Liang, Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda, Proc. Natl. Acad. Sci. USA 107, 18445 (2010).
- [25] V. Kazeev, M. Khammash, M. Nip, and C. Schwab, Direct solution of the chemical master equation using quantized tensor trains, PLoS Comput. Biol. 10, e1003359 (2014).
- [26] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22, 403 (1976).
- [27] M. A. Gibson and J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, J. Phys. Chem. A 104, 1876 (2000).
- [28] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, J. Chem. Phys. 115, 1716 (2001).
- [29] T. Stantcheva and E. Herbst, Deuterium fractionation on interstellar grains studied with the direct master equation approach, Mon. Not. R. Astron. Soc. 340, 983 (2003).
- [30] A. Lipshtat and O. Biham, Efficient Simulations of Gas-Grain Chemistry in Interstellar Clouds, Phys. Rev. Lett. 93, 170601 (2004).

- [31] A. Ale, P. Kirk, and M. P. Stumpf, A general moment expansion method for stochastic kinetic models, J. Chem. Phys. 138, 174101 (2013).
- [32] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Delayinduced stochastic oscillations in gene regulation, Proc. Natl. Acad. Sci. USA 102, 14593 (2005).
- [33] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, Proc. Natl. Acad. Sci. USA 99, 12795 (2002).
- [34] O. Bénichou, C. Chevalier, J. Klafter, B. Meyer, and R. Voituriez, Geometry-controlled kinetics, Nat. Chem. 2, 472 (2010).
- [35] A. Ovchinnikov and Y. B. Zeldovich, Role of density fluctuations in bimolecular reaction kinetics, Chem. Phys. 28, 215 (1978).
- [36] T. Brett and T. Galla, Stochastic Processes with Distributed Delays: Chemical Langevin Equation and Linear-Noise Approximation, Phys. Rev. Lett. 110, 250601 (2013).
- [37] V. M. Kenkre, E. W. Montroll, and M. F. Shlesinger, Generalized master equations for continuous-time random walks, J. Stat. Phys. 9, 45 (1973).
- [38] J. C. Jaeger and G. Newstead, An Introduction to the Laplace Transformation with Engineering Applications (Methuen, London, UK, 1969).
- [39] H. H. Qiu, B. G. Zhang, and T. S. Zhou, Explicit effect of stochastic reaction delay on gene expression, Phys. Rev. E 101, 012405 (2020).
- [40] J. Elf and M. Ehrenberg, Fast evaluation of fluctuations in biochemical networks with the linear noise approximation, Genome Res. 13, 2475 (2003).
- [41] W. D. Donachie and A. Robinson, Cell division: Parameter values and the process, In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, edited by F. Neidhardt *et al.* (ASM Press, Washington, DC, 1987), pp. 1578–1593
- [42] J. Paulsson and M. Ehrenberg, Noise in a minimal regulatory network: Plasmid copy number control, Q. Rev. Biophys. 34, 1 (2001).
- [43] J. Elf, J. Paulsson, O. G. Berg, and M. Ehrenberg, Near-critical

phenomena in intracellular metabolite pools, Biophys. J. 84, 154 (2003).

- [44] A. S. Novozhilov, G. P. Karev, and E. V. Koonin, Biological applications of the theory of birth-and-death processes, Bioinform. 7, 70 (2006).
- [45] B. C. Arnold, Majorization: Here, there and everywhere, Stat. Sci. 22, 407 (2007).
- [46] T. S. Gardner, C. R. Cantor, and J. J. Collins, Construction of a genetic toggle switch in Escherichia coli, Nature (London) 403, 339 (2000).
- [47] A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, Genetic Toggle Switch without Cooperative Binding, Phys. Rev. Lett. 96, 188101 (2006).
- [48] T. Biancalani and M. Assaf, Genetic Toggle Switch in the Absence of Cooperative Binding: Exact Results, Phys. Rev. Lett. 115, 208101 (2015).
- [49] A. Goldbeter and D. E. Koshland Jr., An amplified sensitivity arising from covalent modification in biological systems, Proc. Natl. Acad. Sci. 78, 6840 (1981).
- [50] O. G. Berg, J. Paulsson, and M. Ehrenberg, Fluctuations and quality of control in biological cells: Zero-order ultrasensitivity reinvestigated, Biophys. J. 79, 1228 (2000).
- [51] B. Franzke and B. Kosko, Noise can speed convergence in Markov chains, Phys. Rev. E 84, 041112 (2011).
- [52] B. Franzke and B. Kosko, Noise can speed Markov chain Monte Carlo estimation and quantum annealing, Phys. Rev. E 100, 053309 (2019).
- [53] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, Phy. Rep. 716–717, 1 (2017).
- [54] A. O. Allen, Probability, Statistics, and Queueing Theory: With Computer Science Applications, 2nd ed. (Academic Press, Boston, MA, 1990).
- [55] T. Jia and R. V. Kulkarni, Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting, Phys. Rev. Lett. **106**, 058102 (2011).
- [56] H. Risken, *The Fokker-Planck Equation* (Spinger Verlag, Berlin, 1984).
- [57] C. Gardiner, Handbook of Stochastic Methods (Springer Verlag, Berlin, 1985).