ANALYSIS OF NON-MARKOVIAN EFFECTS IN GENERALIZED BIRTH-DEATH MODELS

ZHENQUAN ZHANG

Key Laboratory of Computational Mathematics, Guangdong Province and School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China

MEILING CHEN, JIAJUN ZHANG AND TIANKOU ZHOU

Key Laboratory of Computational Mathematics, Guangdong Province and School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China

(Communicated by Lei Zhang)

Abstract. Birth-death processes are a fundamental reaction module for which we can find its prototypes in many scientific fields. For such a kind of module, if all the reaction events are Markovian, the reaction kinetics is simple. However, experimentally observable quantities are in general consequences of a series of reactions, implying that the synthesis of a macromolecule in general involve multiple middle reaction steps with some reactions that would not be specified by experiments. This multistep process can create molecular memory between reaction events, leading to non-Markovian behavior. Based on the theoretical framework established in a recent paper published in [39], we find that the effect of non-Markovianity is equivalent to the introduction of a feedback, non-Markovianity always amplifies the mean level of the product if the death reaction is non-Markovian but always reduces the mean level if the birth reaction is non-Markovian, and in contrast to Markovianity, non-Markovianity can reduce or amplify the product noise, depending on the details of waiting-time distributions characterizing reaction events. Examples analysis indicates that non-Markovianity, whose effects were neglected in previous studies, can significantly impact gene expression.

1. Introduction. Stochastic reaction networks appear in many scientific contexts, including biology, chemistry, ecology, epidemics and neural science [10, 24, 36]. Modeling these systems has long relied on the Markovian assumption, that is, the probabilities that future reaction events happen depend, only, on the present state of the system, independent of the prior history. This memoryless property implies that the Markovian reaction kinetics can be described by Poissonian processes with constant rates, which are characterized by exponential waiting-time distributions [10]. However, many reaction processes including birth-death processes take place in a memory manner or are non-Markovian. For example, incomplete mixing of reactions or small reaction steps involved in the synthesis of a macromolecule (e.g., mRNA or protein) can create molecular memory [3, 4, 5, 15, 17, 20, 26, 30, 39], implying that waiting times between reaction events follow non-exponential

2020 Mathematics Subject Classification. Primary: 60G07, 60J27; Secondary: 92B05.

Key words and phrases. Birth-death process, non-Markovianity, waiting-time distribution.

* Corresponding author: Tianshou Zhou.
distributions. Non-Markovianity has also been verified by the increasing availability of time-resolved data on different kinds of interactions [2, 6, 13, 14, 28, 33, 34].

As one of the most common and fundamental reaction modules, birth-death processes, with some straightforward additions such as innovation, are a simple, natural and formal framework for modeling a vast variety of biological processes such as population dynamics, speciation, genome evolution [10, 23, 36]. Such a simple reaction module was before studied under the Markovian hypothesis. The main results obtained include that the number of stationary molecules for the reactive species follows a Poissonian distribution [36] and the corresponding Fano factor (defined as the ratio of variance over mean) is equal to unit. These results well characterize the reaction kinetics of birth-death processes without molecular memory. However, the waiting times between reaction events may follow non-exponential distributions characterizing non-Markovianity with reasons stated above, raising issues: How does the non-Markovianity affect reaction kinetics? Whether does the non-Markovianity have extra effects in contrast to Markovianity? Addressing these issues are not non-trivial even for non-Markovian birth-death processes with simple reaction network structure since non-Markovianity can induce additional dynamics [39].

In this paper, we use the theoretical framework established in our recent paper [39] to analyze reaction kinetics of a generalized birth-death process, focusing on explicit effects of molecular memory. Here by generalized we mean that waiting times for birth and death follow general (exponential or non-exponential) distributions. Recall that the core of that theoretical framework is a stationary generalized chemical master equation (sgCME) [39], which is established by introducing an effective transition rate for each reaction in the underlying reaction network (also see ref. [39]). Based on the sgCME for the generalized birth-death process, we first derive analytical formulas for calculating stationary distribution and statistical indices (such as noise intensity defined as the ratio of the variance over the squared mean). Then, to obtain explicit effects of non-Markovianity, we consider two special cases of waiting-time distributions: exponential and Gamma (in particular, Erlang) distributions [9]. As such, we analytically show that the effect of non-Markovianity is equivalent to the introduction of feedback, and non-Markovianity can tune both the mean level of the product and the noise in the product. More interestingly, we numerically find that for a generalized process of gene expression, non-Markovianity can induce bimodality. Our study not only reveals nontrivial effects of non-Markovianity but also would provide inspiration for studying more complex (chemical or biochemical) reaction processes on networks.

2. Preliminary. This section is introduced to help the reader better understand the results to be obtained in the next sections.

2.1. Biophysical foundation of waiting-time distributions. Exponential waiting-time distribution: While exponential waiting times between reaction events are a pillar of the classical chemical master equation (CME), exponential waiting-time distributions can characterize the times between individual events in a Poisson point process. These distributions can be derived in different ways, but an approximate analysis of molecular collision times can provide a clearer physical explanation.

Assume that all molecules are spheres with the diameter denoted by $D$. When the center of one molecule gets that of another molecule within a distance $D$, a collision will occur. More precisely, a molecule will sweep out a volume equal
to $\pi D v dt$ in the next small time $dt$ in the sense that if another molecule center falls within this volume, a collision will occur in the time interval $(t, t + dt)$, where $v$ represents the average speed. Note that $N$ molecules sweep out this volume for $N$ times. For a given molecular center, the probability that a molecule falls within the total volume swept out by another molecule can be approximated as $\pi N D^2 v dt/V$ if $N$ is sufficiently large, where $V$ represents the system’s volume. The probability of no collision for a given molecule in time $dt$ is given by $P_1 = 1 - \pi N D^2 v dt/V[11]$. The probability of no collision after $N$ time intervals, $N dt$, is given by $P_2 = (1 - \pi N D^2 v dt/V)^N \sim \exp\left(-\pi N D^2 v t/V\right)$, and the probability of a collision is $1 - P_2$, where $t = N dt$. Thus, the probability density function (PDF) of time $t$ can be expressed as

$$
\psi_{\text{exponential}}(t) = \mu e^{-\mu t}
$$

with $\mu = N \pi D^2 v / V$. This is an exponential distribution, which can be used in the modeling of the waiting times between single elementary reaction steps. We point out that this derivation is based on the Markovian assumption, and the exponential waiting-time distribution can also be derived from a master equation [10, 36].

**Erlang waiting-time distribution**: In contrast to the times until the next molecular event happens in single-step processes (elementary reactions), the waiting times between reaction events in multistep processes are, in general, not exponentially distributed. If a reaction process goes through a series of (identical) exponential steps, the waiting times will be Gamma distributed [9]. Mathematically, this process can be regarded as a sum of $k$ single-step processes, so the overall waiting-time distribution is the $k$-fold convolution of exponential distributions, that is,

$$
\psi_{\text{gamma}}(t) = \left[\mu e^{-\mu t} \ast \mu e^{-\mu t} \ast \ldots \ast \mu e^{-\mu t}\right]_{k \text{ times}} = \frac{t^{k-1} e^{-\mu t} \mu^k}{\Gamma(k)},
$$

where symbol $\ast$ denotes convolution, and $\Gamma(k)$ is the common Gamma function. Function $\psi_{\text{gamma}}(t)$ is known as a gamma distribution with shape parameter $k$ and rate parameter $\mu$. If $k$ is a positive integer, the distribution is an Erlang one, which is the interest of this paper. It is worth pointing out that $k$ may represent the number of small reaction steps involved in the synthesis process of a macromolecule.

### 2.2. Representations of birth-death processes

Here, we simply describe several representations of birth-death processes with the purpose to help the reader’s understanding and to be convenient in the later use. In general, birth-death processes can be classified as Markovian birth-death processes and non-Markovian birth-death processes, depending on waiting-time distributions, referring to Fig. 1.

First, let us recall that the common birth-death process is described by two reactions: $\emptyset \xrightarrow{g} X$ and $X \xrightarrow{d} \emptyset$, where $X$ represents reactive species and is a stochastic variable taking non-negative integers or real numbers. Two parameters $g$ and $d$, which are called reaction rates and assumed to be constants, represent synthesis (birth) and degradation (death) rates, respectively. Second, we introduce another description of the above birth-death process by introducing so-called inter-event (or intrinsic) waiting times [1]. Let $\tau_g$ and $\tau_d$ represent intrinsic waiting times for birth and death reactions, respectively. Then, according to Ref. [1], we know that $\tau_g$ and $\tau_d$ follow the distributions given by $\psi_1(t) = ge^{-gt}$ (hereafter symbol $t$ represents time and $t \geq 0$) and $\psi_2(t; n) = dne^{-dnt}$ respectively, where $n$
Waiting-time distributions

Exponential waiting times
Non-exponential waiting times

Markovian Model
No-Markovian Model

Figure 1. Schematic diagram for birth-death reaction process (A), where $\psi_1(t)$ and $\psi_2(t; n)$ are interreaction waiting-time distributions, which may be exponential (B) or non-exponential (C). Here $n$ represents the number of molecules of reactive species $X$.

represents the number of molecules for $X$. Thus, the above birth-death process can be described as $\emptyset \xrightarrow{\psi_1(t)} X$ and $X \xrightarrow{\psi_2(t; n)} \emptyset$, referring to Fig. 1(B).

Third, as mentioned in the introduction, the incomplete mixing of reactions or/and middle reactions involved in the synthesis of a macromolecule can create molecular memory [3, 4, 5, 15, 17, 20, 26, 30, 39], implying that intrinsic waiting-time distributions for birth and death reactions may be non-exponential, i.e., $\psi_1(t)$ or $\psi_2(t; n)$ may be non-exponential, referring to Fig. 1(C). In this case, the corresponding reaction events happen in a non-Markovian manner.

Finally, we point out that the above birth-death process without molecular memory can also be represented in terms of reaction propensity functions as $\emptyset \xrightarrow{a_1(n)} X$ and $X \xrightarrow{a_2(n)} \emptyset$, where $a_1(n) = g$ (independent of $n$), and $a_2(n) = dn$. This representation is convenient for writing deterministic equations directly based on given reactions, e.g., the ordinary differential equation for this reaction system is given by $dx/dt = a_1(x) - a_2(x)$, where $x$ represents the concentration of $X$.

For this Markovian reaction module, there are apparently the following relationships between intrinsic waiting-time distributions and reaction propensity functions: $\psi_1(t) = a_1(n) e^{-a_1(n)t}$ and $\psi_2(t; n) = a_2(n) e^{-a_2(n)t}$.

We emphasize that the above descriptions can easily be extended to more general reaction processes on networks.

3. General results.

3.1. Stationary generalized chemical master equation. The theoretical framework proposed in ref. [39] is quite formal, inconvenient to the intuitive understanding of effects of non-Markovianity on reaction kinetics of generalized birth-death processes. In order to help the readers better understand of our results and for the completeness of this paper, we simply introduce the sgCME for a birth-death
process with arbitrary waiting-time distributions, which is an extension of the com-
mon stationary CME for a Markovian birth-death processes and is very useful for
studying probabilistic behavior of non-Markovian birth-death processes. For this,
it is needed to understand two effective transition rates for birth and death reac-
tions, which are extensions of common reaction propensity functions in the case of
Markovian reactions.

Consider a generalized birth-death process described by \( \emptyset \xrightarrow{\psi_1(t)} X \) and \( X \xrightarrow{\psi_2(t; n)} \emptyset \),
where \( \psi_1(t) \) and \( \psi_2(t; n) \) are two general internal-event waiting-time distributions
for birth and death reactions, respectively. Let \( M_i(t; n) \) (\( i = 1, 2 \)) be memory
functions respectively for birth and death reactions, and \( \tilde{M}_i(s; n) \) be the Laplace
transform of \( M_i(t; n) \), \( i = 1, 2 \). Then, according to ref. [39], we can establish the
following theorem, which is useful in our analysis.

**Theorem 3.1.** For arbitrary intrinsic waiting-time distributions \( \psi_1(t) \) and \( \psi_2(t; n) \),
the limit \( \lim_{s \to 0} \tilde{M}_i(s; n) \) always exists and is finite, where \( i = 1, 2 \). Moreover, if this
limit is denoted by \( K_i(n) \), then

\[
K_1(n) = \frac{\int_0^{+\infty} \left[ \psi_1(t) \int_t^{+\infty} \psi_2(t'; n) \, dt' \right] \, dt}{\int_0^{+\infty} \left[ \int_t^{+\infty} \psi_1(t') \, dt' \right] \int_t^{+\infty} \psi_2(t'; n) \, dt' \, dt}
\]

(3a)

\[
K_2(n) = \frac{\int_0^{+\infty} \left[ \psi_2(t; n) \int_t^{+\infty} \psi_1(t') \, dt' \right] \, dt}{\int_0^{+\infty} \left[ \int_t^{+\infty} \psi_1(t') \, dt' \right] \int_t^{+\infty} \psi_2(t'; n) \, dt' \, dt}
\]

(3b)

Two functions \( K_1(n) \) and \( K_2(n) \) will be called effective transition rates respectively
for birth and death reactions throughout this paper. In order to help the
reader understand these two rates, let us consider the common birth-death process
where the intrinsic-event waiting-time distributions are exponential (see subsection
2.2 above). In this case, simple calculations show \( K_1(n) = g \) and \( K_2(n) = dn \),
impling that the effective transition rates are exactly the same as the reaction
propensity functions. For a more general case, effective transition rates are exten-
sions of common reaction propensity functions.

Next, according to ref [39] again, we can establish another theorem.

**Theorem 3.2.** Assume that the stationary distribution (denoted by \( P(n) \)) exists.
Then, \( P(n) \) satisfies the following sgCME:

\[
(E^{-1} - 1) [K_1(n) P(n)] + (E - 1) [K_2(n) P(n)] = 0.
\]

(4)

In order to help the reader understand the implication of Eq.(4), we construct
a new birth-death reaction process described by: \( \emptyset \xrightarrow{K_1(n)} X \) and \( X \xrightarrow{K_2(n)} \emptyset \),
where \( K_1(n) \) and \( K_2(n) \) should be understood as reaction propensity functions. Then,
the corresponding CME reads

\[
\frac{\partial}{\partial t} Q(n; t) = \left( E^{-1} - 1 \right) [K_1(n) Q(n; t)] + (E - 1) [K_2(n) Q(n; t)].
\]

(5)

The steady-state equation of this equation is exactly the same as Eq.(4). In this
sense, we say that the original non-Markovian issue is transformed into a Markovian
one.

However, it should be pointed out that there would be differences in dynamic
probabilistic behavior between non-Markovian and Markovian cases. For example,
consider the above generalized birth-death process wherein \( X \) is assumed to
represent protein. Two waiting-time distributions are assumed to take the following forms: 

$$\psi_1(t) = \frac{\lambda_1^{k_1}k_1!}{(k_1)!} t^{k_1-1}e^{-\lambda_1 t},$$

and

$$\psi_2(t; n) = \lambda_2 ne^{-\lambda_2 nt}.$$ 

Numerical results demonstrated in Fig. 2 show that at initial stage, the difference between two dynamic distributions is indeed large but with the time evolution, the difference gradually reduces and finally disappears. This figure also numerically verifies that the stationary distribution in the generalized birth-death reaction system is existent. Therefore, the assumption made in Theorem 2 is reasonable although its mathematical proof seems difficult.

**Figure 2.** Difference between dynamic probability distributions obtained in the original non-Markovian model (curves with empty circles) and in the constructed Markovian model (colored curves), both being used in the modeling of a gene transcription process. Parameter values are set as $k_1 = 3, \lambda_1 = 90, \lambda_2 = 1$, and the initial conditions are set as the same in two cases. In the diagram, ‘$t = \text{inf}$’ means that the distributions after $t > 10$ are approximately the same, implying that the stationary distribution exists.

By Theorems 1 and 2, we can obtain the following two useful corollaries.

**Corollary 1.** If $\psi_1(t) = \frac{(\lambda_1)^{k_1}}{t!(k_1)!} k_1! t^{k_1-1}e^{-\lambda_1 t}$, $\psi_2(t; n) = \frac{(\lambda_2)^{k_2}}{t!(k_2)!} k_2! t^{k_2-1}e^{-\lambda_2 nt}$, where $k_1$ and $k_2$ are positive integers, then the corresponding effective transition rates can analytically be expressed as

$$K_1(n) = \frac{(\lambda_1)^{k_1} \sum_{i=0}^{k_2-1} \binom{k_1 + i - 1}{i} \frac{(n\lambda_2)^i}{(\lambda_1 + n\lambda_2)^{i+1}}}{\sum_{i=0}^{k_2-1} \sum_{j=0}^{k_2-1} \binom{i+j}{i} \frac{(\lambda_1)^i(n\lambda_2)^j}{(\lambda_1 + n\lambda_2)^{i+j+1}}}$$

and

$$K_2(n) = \frac{(\lambda_2)^{k_2} \sum_{i=0}^{k_1-1} \binom{k_2 + i - 1}{i} \frac{(\lambda_1)^i(n\lambda_2)^j}{(\lambda_1 + n\lambda_2)^{i+j+1}}}{\sum_{i=0}^{k_1-1} \sum_{j=0}^{k_2-1} \binom{i+j}{i} \frac{(\lambda_1)^i(n\lambda_2)^j}{(\lambda_1 + n\lambda_2)^{i+j+1}}}.$$
where \( n = 0, 1, 2, \cdots \), and we define \( \binom{i+j}{i} = \frac{(i+j)!}{i!j!} \).

Proof. In fact, by calculations, we can respectively show

\[
\int_0^t \psi_1(t') \, dt' = e^{-\lambda_1 t} \sum_{j=0}^{k_1-1} \frac{1}{j!} (t\lambda_1)^j,
\]

\[
\int_0^t \psi_2(t'; n) \, dt' = e^{-n\lambda_2 t} \sum_{j=0}^{k_2-1} \frac{1}{j!} (nt\lambda_2)^j,
\]

\[
\int_0^\infty \left[ \int_0^t \psi_1(t') \, dt' \right] \left[ \int_0^t \psi_2(t'; n) \, dt' \right] \, dt = \sum_{i=0}^{k_1-1} \sum_{j=0}^{k_2-1} \frac{(i+j)!}{i!j!} (\lambda_1)^i (n\lambda_2)^j \frac{(\lambda_1 + n\lambda_2)^{i+j+1}}{\Gamma(k_1 + i)}.
\]

According to Corollary 1, we can show

\[
K_1(n) = \frac{n\lambda_2 \lambda_1^{k_1}}{(\lambda_1 + n\lambda_2)^{k_1} - \lambda_1^{k_1}} = \frac{n\lambda_2 \lambda_1^{k_1}}{(\lambda + n)^{k_1} - \lambda^{k_1}},
\]

with \( K_1(0) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \), \( K_2(n) = n\lambda_2 \), where \( n \) represents the number of the X molecules, and \( \lambda = \lambda_1/\lambda_2 \) represents the ratio of the mean synthesis rate over the mean degradation rate. Note that \( k_1 > 1 \) implies the existence of non-Markovianity. Using \( K_1(n) \) and \( K_2(n) \), we construct a new birth-death process described by \( \emptyset \xrightarrow{K_1(n)} X \xrightarrow{K_2(n)} \emptyset \), where \( K_1(n) \) and \( K_2(n) \) should be understood as reaction propensity functions for birth and death reactions, respectively. For this constructed reaction system, we know that if \( K_1(n) \) is a rigorously monotonic function of \( n \), then this implies the existence of a feedback. On the other hand, we calculate the derivative of function \( K_1(x) \):

\[
\frac{dK_1(x)}{dx} = -\lambda_2 \lambda_1^{k_1} (k_1 - 1) (\lambda + x)^{k_1 - 1} - \lambda k_1 (\lambda + x)^{k_1 - 1} + \lambda k_1.
\]

If we set \( f(x) = (k_1 - 1) (\lambda + x)^{k_1 - 1} - \lambda k_1 (\lambda + x)^{k_1 - 1} + \lambda k_1 \), then \( f'(x) = xk_1 (k_1 - 1) (\lambda + x)^{k_1 - 2} > 0 \) for all \( x > 0 \), and \( f(0) = 0 \). Therefore, \( dK_1(x)/dx < 0 \) for all \( x > 0 \). In combination with Theorem 2, Corollary 2 is thus proven. \( \square \)
3.2. **Stationary distributions and statistical indices.** Although birth-death processes are simple in structure, it seems difficult to give analytical distributions in the non-Markovian case. Here we derive the formal expression of the stationary distribution for a general (Markovian or non-Markovian) birth-death process. Note that Eq. (4) can be rewritten as

\[(E - I) \{ K_2 (n) P (n) - E^{-1} K_1 (n) P (n) \} = 0,\]

which implies

\[K_2 (n) P (n) - E^{-1} K_1 (n) P (n) = -J,\]

where constant \(-J\) represents the net probability flow from \(n\) to \(n - 1\). Setting \(n = 0\) gives \(J = 0\). Therefore,

\[K_2 (n) P (n) = K_1 (n - 1) P (n - 1).\]

Thus, the stationary distribution is formally given by

\[P (n) = P (0) \prod_{i=1}^{n} \frac{K_1 (i - 1)}{K_2 (i)}, \quad n = 1, 2, \ldots, \tag{8}\]

where \(K_1 (i)\) and \(K_2 (i)\) are calculated according to Eqs. (3a) and (3b). In Eq. (8), \(P (0)\) is determined by the conservative condition of probability, given by

\[P (0) = \left[1 + \sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)\right]^{-1}. \tag{9}\]

We observe from the above analysis that non-Markovianity affects the stationary distribution in a nonlinear manner since \(K_1 (n)\) or/and \(K_2 (n)\) are nonlinear functions of \(n\) in the non-Markovian case. See the next section for more details.

Furthermore, the expectation of the stationary distribution is given by

\[\langle n \rangle = \sum_{n=0}^{\infty} n P (n) = \frac{\sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)}{1 + \sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)}, \tag{10}\]

and the noise intensity for \(X\), denoted by \(\eta^2\), which is defined as the ratio of variance over the squared mean, is calculated according to

\[\eta^2 = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle^2} = \frac{\sum_{n=1}^{\infty} n^2 \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)}{\left[\sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)\right]^2} + 1 - \frac{\left[\sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)\right]^2}{\sum_{n=1}^{\infty} n \prod_{i=1}^{n} K_1 (i - 1) / K_2 (i)}. \tag{11}\]

Formulas (8)-(11) are general and can be applied to any birth-death processes with general waiting-time distributions. In particular, formula (8) indicates that sgCMEs for non-Markovian systems are analytically solvable. In spite of this, the biological implication of non-Markovianity is not clear. In the next section, we will consider special cases to specify formulas (8)-(11), and give some biological interpretations for resulting expressions.
4. **Analytical results in special cases.** In this section, we will derive analytical results by considering special waiting-time distributions such as exponential distributions and Erlang distributions. In order to obtain explicit stationary distributions and statistical quantities, the key is to specify the expressions of two functions $K_1(n)$ and $K_2(n)$, which are in principle calculated according to Eqs. (3a) and (3b).

Assume that the intrinsic waiting times for birth and death reactions follow Erlang distributions $\psi_1(t) = (\lambda_1 t)^{k_1-1} e^{-\lambda_1 t}$, $\psi_2(t; n) = (n\lambda_2 t)^{k_2-1} e^{-n\lambda_2 t}$, where $k_1$ and $k_2$ are two positive integers. In the following analysis, one should note that: (I) the case that $k_1$ or $k_2$ is a positive integer of greater than 1 corresponds to a non-Markovian birth-death process whereas the case of $k_1 = k_2 = 1$ corresponds to a Markovian birth-death process or to exponential waiting-time distributions; (II) the larger the $k_1$ or $k_2$ is, the stronger is the non-Markovianity; (III) $k_1$ and $k_2$ can represent (or model) the number of reaction steps involved in the birth and death reactions; and (IV) $K_1(n)$ and $K_2(n)$ are given by Corollary 1. In addition, the stationary distribution and the noise intensity are calculated according to Eq. (8) and Eq. (11), respectively.

Since the biological implication of non-Markovianity cannot be seen in the general case of $k_1$ or $k_2$, we distinguish three special cases to give analytical results and their interpretations.

**Case 1:** $k_1 = k_2 = 1$

This case corresponds to exponential waiting times, i.e., both $\psi_1(t)$ and $\psi_2(t; n)$ are exponential distributions. According to Corollary 1, we can know that $K_1(n) = \lambda_1$ and $K_2(n) = n\lambda_2$, respectively. According to Eq. (8) combined with Eq. (9), the stationary distribution for $X$ is given by

$$P(n) = P_0 \frac{1}{n!} \left( \frac{\lambda_1}{\lambda_2} \right)^n = \frac{\lambda^n}{n!} e^{-\lambda},$$

which is a Poisson distribution determined by a single parameter: $\lambda = \lambda_1/\lambda_2$ (which actually represents the ratio of the mean birth rate over the mean death rate). This is a known result for the Markovian birth-death process. Moreover, the mean of $X$ is given by $\langle n \rangle = \lambda$, and the noise intensity by

$$\eta_{\text{Markov}}^2 = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle^2} = \frac{1}{\lambda},$$

which indicates that for the Markovian birth-death process, a larger ratio of the birth rate over the death rate leads to the smaller noise in the product. If we calculate the Fano factor, which is defined as the ratio of the variance over the mean, then this factor is equal to 1, which is a common characteristic of Markovian birth-death processes.

**Case 2:** $k_1 = 1$ and $k_2 > 1$

In this case, $\psi_1(t)$ is an exponential distribution whereas $\psi_2(t; n)$ is an Erlang distribution. According to Corollary 1, we can show

$$K_1(n) = \lambda_1, \quad K_2(n) = \frac{\lambda_1 (\lambda_2 n)^{k_2}}{(\lambda_1 + \lambda_2 n)^{k_2}}.$$
In particular, we have \( K_2 (n) = \lambda_2 n \) if \( k_2 = 1 \) (corresponding to the Markovian case), and \( K_2 (n) = \frac{\lambda_2 n}{\lambda + 2 \lambda_2 n} \) if \( k_2 = 2 \) (corresponding to the Markovian case). More generally, we can prove \( K_2 (n) < \lambda_2 n \) if \( k_2 \geq 2 \). This indicates that the effective transition rate for the degradation reaction in the non-Markovian case is less than the reaction propensity function for the degradation reaction in the corresponding Markovian case.

According to Eq. (8), the stationary distribution is given by

\[
P(n) = P(0) \prod_{i=1}^{n} \frac{(\lambda+i)^{k_2} - i^{k_2}}{k_2}, \quad n = 1, 2, \cdots,
\]

where \( P(0) \) is determined by the conservative condition for probability, given by

\[
P(0) = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \left[ (\lambda+i)^{k_2} - i^{k_2} \right] / i^{k_2}}.
\]

In order to see the effect of non-Markovianity more clearly, let us consider the case of \( k_2 = 2 \) without loss of generality. In this case, the stationary distribution can be expressed as

\[
P(n) = P(0) \frac{(2\lambda)^n ((\lambda + 2) / 2)_n}{n!},
\]

with \( P(0) = \frac{1}{1 \Gamma((\lambda+2)/2, 1; 2\lambda)} \), \( n = 1, 2, \cdots \), where \((c)_n\) is defined as \((c)_n = c(c+1) \cdots (c+n-1)\) (in particular, \((1)_n = n!\) ), whereas function \( \Gamma(a, b; z) \) is defined as \( \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!} \). Then, the mean is given by

\[
\langle n \rangle = \lambda (\lambda + 2) \frac{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}.
\]

Simple calculation shows

\[
\lambda (\lambda + 2) \frac{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}
= \lambda (\lambda + 2) \sum_{n=0}^{\infty} \frac{(1 + (\lambda + 2) / 2)_n (2\lambda)_n}{(2)_n n!}
= \lambda (\lambda + 2) \sum_{n=0}^{\infty} \frac{(1 + (\lambda + 2) / 2)_n ((\lambda + 2) / 2)_n (2\lambda)_n}{(\lambda + 2)_n (n+1)(1)_n n!}
= \lambda \sum_{n=0}^{\infty} \frac{2n + \lambda + 2 ((\lambda + 2) / 2)_n (2\lambda)_n}{n+1 (1)_n n!}
> \lambda \sum_{n=0}^{\infty} \frac{((\lambda + 2) / 2)_n (2\lambda)_n}{(1)_n n!}
= \lambda \frac{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}{\Gamma((\lambda + 2) / 2, 1; 2\lambda)}.
\]

Thus, the inequality \( \langle n \rangle > \lambda \) always holds, implying that non-Markovianity always amplifies the mean level in contrast to Markovianity (remark: the mean equals \( \lambda \) in the Markovian case). The noise intensity is given by

\[
\eta^2_{\text{non-Markov}} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle^2}
= \frac{1}{\lambda (\lambda + 2)} \frac{\Gamma(\frac{\lambda + 2}{2}, 1; 2\lambda) \Gamma(\frac{\lambda + 2}{2}, 1; 2\lambda)}{\Gamma(\frac{\lambda + 2}{2}, 1; 2\lambda) \Gamma(\frac{\lambda + 2}{2}, 2; 2\lambda) \Gamma(\frac{\lambda + 2}{2}, 2; 2\lambda) - 1}.
\]
We can show \( \lim_{\lambda \to 0} \eta_{\text{non-Markov}}^2 = +\infty \) and \( \lim_{\lambda \to +\infty} \eta_{\text{non-Markov}}^2 = 0 \). Moreover, the smaller the ratio \( \lambda \) is, the lower is the noise and the higher otherwise. Therefore, the change tendency in the non-Markovian case is similar to that in the Markovian case.

Next, let us examine the non-Markovian effect by comparing the noise intensities in Markovian and non-Markovian cases. For convenience, we still consider the case of \( k_2 = 2 \). By calculation, we find that the difference between the noise intensities in the Markovian and non-Markovian cases is given by

\[
d_1 = \eta_{\text{non-Markov}}^2 - \eta_{\text{Markov}}^2 = \sum_{n=0}^{\infty} (2\lambda)^n \sum_{i+j=n} \frac{C_{ij}}{n! (2\lambda)^n},
\]

(19)

where \( C_{ij} = ij + j - (2\lambda) i - (\lambda + j) (\lambda + 1) \) with \( j = n - i \). Note that \( C_{ij} \) may be less than zero but may also be more than zero, depending on the value of \( i \) or \( j \). In other words, non-Markovianity may reduce the noise but may also amplify it, depending on parameter \( \lambda \). For a larger \( k_2 \), the qualitative result is similar to the case of \( k_2 = 2 \).

**Case 3: \( k_2 = 1 \) and \( k_1 > 1 \)**

In this case, \( \psi_2(t; n) \) is an exponential distribution and \( \psi_1(t) \) is a Gamma distribution. Moreover, \( K_1(n) \) and \( K_2(n) \) are given by Eq. (6).

In order to see the effect of non-Markovianity more clearly, let us consider the case of \( k_1 = 2 \). In this case, \( K_1(n) = \lambda_1 \lambda/(2\lambda + n) \) and \( K_2(n-1)/K_2(n) = \lambda^2/[n(2\lambda + n - 1)] \). According to Eq. (8), we can express the stationary distribution as

\[
P(n) = \frac{1}{\Gamma^2(1, 2\lambda; \lambda^2)} \frac{\lambda^{2n}}{n!}, \quad n = 0, 1, 2, \ldots,
\]

(20)

where we define \((2\lambda)_0 = 1\) and \( \Gamma^2(a, b; c; z) = \sum_{n=0}^{\infty} z^n \frac{(a)_n}{(b)_n (c)_n} \). By calculation, we know that the first-order raw moment is given by

\[
\langle n \rangle = \frac{\lambda_1 F_2(1, 1, 2\lambda + 1; \lambda^2)}{2 \Gamma^2(1, 1, 2\lambda; \lambda^2)}.
\]

(21a)

The inequality \( \frac{\lambda_1 F_2(1, 1, 2\lambda + 1; \lambda^2)}{\Gamma^2(1, 1, 2\lambda; \lambda^2)} < \frac{\lambda_1 F_2(1, 1, 2\lambda; \lambda^2)}{\Gamma^2(1, 1, 2\lambda + 1; \lambda^2)} \) always holds, so the inequality \( \langle n \rangle < \lambda/2 \) also holds, implying that non-Markovianity always reduces the mean of the product in contrast to Markovianity. The second-order raw moment is then given by

\[
\langle n^2 \rangle = \frac{\lambda_1 F_2(2, 1, 2\lambda + 1; \lambda^2)}{2 \Gamma^2(1, 1, 2\lambda; \lambda^2)}.
\]

(21b)

Thus, the noise intensity can be expressed as

\[
\eta_{\text{non-Markov}}^2 = \frac{\lambda_1 F_2(2, 1, 2\lambda + 1; \lambda^2)}{\lambda_1 F_2(1, 1, 2\lambda + 1; \lambda^2)} F_2(1, 1, 2\lambda; \lambda^2) - 1.
\]

(22)

We can show \( \eta_{\text{non-Markov}}^2 > 1 \) if \( \lambda < 1 \) and \( \eta_{\text{non-Markov}}^2 < 1 \) otherwise. In addition, similar to the case of \( k_1 = 1 \) and \( k_2 > 1 \), we can show \( \lim_{\lambda \to 0} \eta_{\text{non-Markov}}^2 = +\infty \) and \( \lim_{\lambda \to +\infty} \eta_{\text{non-Markov}}^2 = 0 \). Moreover, the smaller the ratio \( \lambda \) is, the lower is the noise and the higher otherwise.
The difference between the noise intensities in non-Markovian and Markovian cases is given by
\[ d_2 = \eta_{\text{non-Markov}}^2 - \eta_{\text{Markov}}^2 = \frac{\sum_{n=0}^{\infty} \sum_{i+j=n} \lambda^{2(i+j)} D_{ij} \left( 2 \lambda \right)_{i+1} (2 \lambda)_{j+1}}{\lambda \left[ F_2 (1, 1; 2 \lambda; \lambda^2) \right]^2}, \]
where \( D_{ij} = (4 \lambda - 1) ij + 2 \lambda (3i - j) - 4 \lambda^2 (\lambda + 1) \) with \( j = n - i \). Note that \( D_{ij} \) may be less than zero but may also be greater than zero, depending on the value of \( i \) or \( j \). In other words, non-Markovianity may reduce the noise in product but may also amplify the noise, depending on \( \lambda \). For a larger \( k_1 \), the qualitative result is similar to that in the case of \( k_1 = 2 \).

5. Applications of the general framework to gene systems. As is well known, a gene may be expressed in a constitutive or bursty manner. In this section, we apply the general theoretical framework to two representative models of gene expression, which are extensions of gene models studied in the literature [21]. One will see that non-Markovianity has unnotegable influence on the gene-product level and the noise in the product.

5.1. A generalized model of constitutive gene expression. As a special case of the above generalized birth-death process, here we consider a stochastic model of constitutive gene expression with molecular memory [25, 31, 38], referring to Fig. 3(A), where \( \psi_1 (t) \) and \( \psi_2 (t; n) \) are the intrinsic-event waiting-time distributions for the synthesis and degradation of proteins, respectively. The introduction of molecular memory is because the protein as a macromolecule is generated not in a single-step manner but in a multistep manner, with reasons as mentioned in the introduction. We will address the question of how molecular memory affects the expression level and noise of protein.

Notice that the distribution of general form, \( \psi_1 (t) \), can model the complex process of transcription [14, 34]. We assume that \( \psi_1 (t) \) and \( \psi_2 (t; n) \) follow Erlang distributions, that is, \( \psi_1 (t) = (\frac{c \lambda_b}{k_b})^b t^{k_b-1} e^{-c \lambda_b t} \) and \( \psi_2 (t; n) = (\frac{n \lambda_d}{k_d})^d t^{k_d-1} e^{-n \lambda_d t} \), where \( c \) represents the number of DNA molecules, which is assumed to be a constant, \( \lambda_b \) and \( \lambda_d \) are positive constants, representing the mean transcription (or translation) and degradation rates, respectively. Then, two effective transition rates \( K_1 (n) \) and \( K_2 (n) \) are given by Eq. (6) if \( \lambda_1 \) is replaced with \( c \lambda_b \) in Eq. (6).

Simultaneously, the stationary protein distribution is given by Eq. (8). If \( k_d = 1 \) and \( k_b \) is a positive integer of more than 1, then \( K_1 (n) = \frac{n \lambda_d (c \lambda_b)^b}{(c \lambda_b + n \lambda_d)^{b-1} - (c \lambda_b)^b} \), with \( K_1 (0) = \frac{c \lambda_b^b}{k_b} \) and \( K_2 (n) = n \lambda_d \). In this case, the effect of non-Markovianity is equivalent to the introduction of a negative feedback, as interpreted in the case of \( k_2 = 1 \) and \( k_1 > 1 \) above. Moreover, non-Markovianity always amplifies the mean expression level of the gene in contrast to Markovianity. If \( k_b = 1 \) and \( k_d \) is a positive integer of more than 1, then we can show \( K_2 (n) = \frac{n \lambda_d (c \lambda_b)^b}{(c \lambda_b + n \lambda_d)^{b-1} - (c \lambda_b)^b} \) and \( K_1 (n) = c \lambda_b \). This case actually belongs to the case of \( k_1 = 1 \) and \( k_2 > 1 \) discussed above, so non-Markovianity always amplifies the mean expression level of the protein in contrast to Markovianity.

Without loss of generality, we set \( c = 1 \). Fig. 3(B,C) shows numerical results, where two special cases are considered: the source of on-Markovianity comes from the birth process, referring to Fig. 3(B); the source of on-Markovianity originates
in the death process, referring to Fig. 3(C). From Fig. 3(B), we observe that with the increase of the common factor that simultaneously enlarges both $k_b$ and $\lambda_b$, the stationary protein distribution shifts toward the left side. Similarly, with the increase of the common factor that simultaneously enlarges both $k_d$ and $\lambda_d$, the stationary protein distribution shifts toward the right side, referring to Fig. 3(C).

In a word, non-Markovianity can impact the stationary protein distribution.

5.2. A generalized model of bursty gene expression. In contrast to the constitutive manner of gene expression, the bursty manner is more common in eukaryotic cells [34]. Consider an extended version of the common two-state model of gene expression, where the number of DNA molecules is assumed to be a constant (it can be set as unit without loss of generality). The extended model is assumed to contain four reactions, denoted by $R_i$ ($1 \leq i \leq 4$), which correspond respectively to off $\psi_1(t; n)$ on, on $\psi_2(t; n)$ off, on $\psi_3(t; n)$ on + $B \cdot$ Protein and Protein $\psi_4(t; n)$ ∅, referring to Fig. 4(A). Here, $\psi_i(t; n)$ ($1 \leq i \leq 4$) are waiting-time distributions, $n$ represents the number of protein molecules, and $B$ represents the burst size and is assumed to follow a geometric distribution of the form $P_b (B = i) = b^i / (1 + b)^{i+1}$ with $i = 0, 1, 2, \cdots$, where $b$ represents the mean burst size. The system has no feedback regulation if $\psi_i(t; n)$ ($1 \leq i \leq 3$) do not depend on $n$, and has feedback regulation otherwise. We will consider the case of no feedback regulation.

Apparently, the extended gene model is different, in structure, from the analyzed-above gene model. Therefore, the above results, including Theorems 1 and 2, cannot be directly applied. However, since the considered gene model essentially belongs to a birth-death process, we want to know how non-Markovianity influences gene expression in the case of bursty expression. In addition, the following investigation is for completeness of results on non-Markovian effects in common gene models.
According to ref. [39], we can derive the following sgCME for the above model of gene expression with general waiting-time distributions:

\[-K_1(n)P_0(n) + K_2(n)P_1(n) + K_4(n + 1)P_0(n + 1) - K_4(n)P_0(n) = 0,\]
\[K_1(n)P_0(n) - K_2(n)P_1(n) + K_4(n + 1)P_1(n + 1) - K_4(n)P_1(n)\]
\[+ \sum_{i=0}^{n} P_b(B = i)K_3(n - i)P_1(n - i) - P_b(B \geq 1)K_3(n)P_1(n) = 0,\] (24)

where \(P_0(n)\) and \(P_1(n)\) are stationary probabilities that the gene is in OFF and ON states respectively, and \(K_i(n)\) is called the effective transition rate for reaction \(R_i\), \(1 \leq i \leq 4\). For simplicity, we consider only the case of \(\psi_4(t; n) = n\delta e^{-n\delta t}\) with \(\delta\) being a positive constant (representing the mean degradation), i.e., we only consider linear degradation. Note that if the gene is in OFF state, reactions \(R_2\) and \(R_3\) will not happen, implying that \(\psi_2(t; n) = \psi_3(t; n) = 0\), and if the gene is in ON state, reaction \(R_1\) will not happen, implying that \(\psi_1(t; n) = 0\). Thus, if we define \(\tilde{\psi}_1(t; n) = \psi_1(t; n), \tilde{\psi}_k(t; n) = 0\) \((k = 2, 3)\) when calculating \(K_1(n)\), and \(\tilde{\psi}_1(t; n) = 0, \tilde{\psi}_i(t; n) = \psi_i(t; n)\) \((i = 2, 3)\) when calculating \(K_2(n)\) or \(K_3(n)\), then according to ref. [39], we can derive the following expressions of \(K_i(n)\) \((1 \leq i \leq 4)\)

\[K_i(n) = \frac{\int_0^{+\infty} \tilde{\psi}_i(t; n) \left[ \prod_{j \neq i} \int_t^{\infty} \tilde{\psi}_j(t'; n) \, dt' \right] \, dt}{\int_0^{+\infty} \left[ \prod_{j=1}^{4} \int_t^{\infty} \tilde{\psi}_j(t'; n) \, dt' \right] \, dt}, \quad 1 \leq i \leq 3,\] (25)

Our purpose is to find the total stationary probability, \(P(n) = P_0(n) + P_1(n)\), based on Eq. (24). It should be noted that formulas Eq. (8) above cannot be directly used to the case of bursty gene expression.

For convenience, we normalize \(K_i(n)\) by \(\delta\), i.e., we set \(\bar{K}_i(n) = K_i(n) / \delta\), where \(i = 1, 2, 3\). Then, Eq. (24) can be rewritten as

\[\begin{align*}
-K_1(n)P_0(n) + \bar{K}_2(n)P_1(n) + (n + 1)P_0(n + 1) - nP_0(n) = 0, \\
\bar{K}_1(n)P_0(n) - \bar{K}_2(n)P_1(n) + (n + 1)P_1(n + 1) - nP_1(n) \\
+ \sum_{i=0}^{n} P_b(B = i)\bar{K}_3(n - i)P_1(n - i) - P_b(B \geq 1)\bar{K}_3(n)P_1(n) = 0,
\end{align*}\] (26)

where \(n = 0, 1, 2, \ldots\) and we define \(P_1(-1) = 0\). In general, solving Eq. (26) is difficult if waiting-time distributions are general. However, Eq. (26) can be numerically solved.

In numerical simulation, we set

\[\psi_1(t; n) = \frac{(\lambda_{on})^{k_{on}}}{\Gamma(k_{on})} t^{k_{on}-1} e^{-\lambda_{on}t},\]
\[\psi_2(t; n) = \frac{(\lambda_{off})^{k_{off}}}{\Gamma(k_{off})} t^{k_{off}-1} e^{-\lambda_{off}t},\]
\[\psi_3(t; n) = \mu e^{-\mu t}.\]

For clarity, we consider two cases: Markovity (corresponding to \(k_{on} = k_{off} = 1\) and non-Markovity (corresponding to the case that at least one of \(k_{on}\), and \(k_{off}\) is greater than 1). Furthermore, we distinguish two cases of non-Markovity: \(k_{on} = 1 \text{ and } k_{off} > 1\) (referring to Fig. 4(B)); \(k_{on} > 1 \text{ and } k_{off} = 1\) (referring to
Fig. 4(C)). Numerical results are demonstrated in Fig. 4(B,C). From these two panels, we observe that for some set of parameter values, although the stationary protein distribution is unimodal in the Markovian case but it is bimodal in the non-Markovian case, implying that non-Markovity can induce bimodality (different from the case of molecular noise-induced phenotypic switching [35]). This is an interesting result. Figure 4 indicates that non-Markovity originating from the birth process and from the death process can all induce bimodality. If the birth and death processes are all non-Markovian, then non-Markovity can also induce bimodality (data are not shown).

\[ \psi_b(t; n) \]

\[ \psi_d(t; n) \]

\[ \psi_{on}(t; n) \]

\[ \psi_{off}(t; n) \]

Figure 4. Characteristics of stationary protein distribution in a generalized model of bursty gene expression (A), where solid lines represent theoretical predictions and empty circles represent numerical results obtained by the Gillespie algorithm [11]. Parameter values are set as: (B) \( \mu = 10, \delta = 1, b = 2, k_{on} = 1, \lambda_{on} = 0.5, k_{off} = 1, \lambda_{off} = 0.1 \) for Markovianity, whereas \( k_{on} = 4, \lambda_{on} = 4 \) and the other parameter values are kept unchanged for non-Markovianity; (C) \( \mu = 10, \delta = 1, b = 2, k_{on} = 1, \lambda_{on} = 0.5, k_{off} = 1, \lambda_{off} = 0.1 \) for Markovianity, whereas \( k_{off} = 4, \lambda_{off} = 0.4 \) and the other parameter values are kept unchanged for non-Markovianity. In (B) and (C), we use 1000 realizations to obtain numerical results and error bars to indicate the error ranges of the numerical results.

In the following, we consider a special case, i.e., assume that \( B \equiv 1 \) (i.e., \( B \) is a deterministic variable equal to 1), \( \psi_1(t; n) = \alpha e^{-\alpha t}, \psi_2(t; n) = \beta e^{-\beta t}, \) and \( \psi_3(t; n) = \mu e^{-\mu t}, \) where \( \alpha, \beta \) and \( \mu \) are positive constants, which actually represent the mean transition rate from OFF to ON, the mean transition rate from ON to OFF, and the mean transcription/translation rate, respectively. In this case, which corresponds to the common ON-OFF model of gene expression, we have

\[ K_1(n) = \alpha, K_2(n) = \beta, K_3(n) = \mu, \text{ or } \hat{K}_1(n) = \hat{\alpha}, \hat{K}_2(n) = \hat{\beta}, \hat{K}_3(n) = \hat{\mu}. \]
We can analytically solve the corresponding sgCME. In fact, we have
\[
-\tilde{\alpha}P_0(n) + \tilde{\beta}P_1(n) + (n + 1)P_0(n + 1) - nP_0(n) = 0,
\]
\[
\tilde{\alpha}P_0(n) - \tilde{\beta}P_1(n) + (n + 1)P_1(n + 1) - nP_1(n) + \tilde{\mu}P_1(n - 1) - \tilde{\mu}P_1(n) = 0,
\]
where \( n = 0, 1, 2, \cdots \). Now, we adopt the binomial moment method [40] to solve Eq. (27). Define fractional binomial moments as
\[
b_k^i = \sum_{n \geq k} \binom{n}{k} P_i(n),
\]
where \( i = 0, 1 \). Then, from Eq. (27) we can derive the following binomial moment equations:
\[
-\tilde{\alpha}b_k^0 + \tilde{\beta}b_k^1 - kb_k^0 = 0,
\]
\[
\tilde{\alpha}b_k^0 - \tilde{\beta}b_k^1 - kb_k^1 + \tilde{\mu}b_{k-1}^1 = 0,
\]
where \( k = 0, 1, 2, \cdots \). By solving Eq. (28) in combination with \( b_0^0 + b_0^1 = 1 \), we can obtain
\[
b_k^0 = \frac{\tilde{\beta}\tilde{\mu}}{k!(\tilde{\alpha} + \tilde{\beta})_{k+1}}, b_k^1 = \frac{-\tilde{\mu}}{k!(\tilde{\alpha} + \tilde{\beta})_{k+1}}.
\]
Thus, the total binomial moments \( b_k \), which are defined as \( b_k = b_k^0 + b_k^1 \), are given by
\[
b_k = \frac{\tilde{\mu}^k}{k!(\tilde{\alpha} + \tilde{\beta})_k}, k = 0, 1, 2, \cdots \).
\]
Note that binomial moments can be used in reconstruction of the distribution. Moreover, the reconstruction formula is [40]
\[
P(n) = \sum_{k \geq n} (-1)^{k-n} \binom{k}{n} b_k,
\]
where \( n = 0, 1, 2, \cdots \). Therefore, the explicit expression of the stationary protein distribution can be expressed as
\[
P(n) = \frac{\tilde{\mu}^n}{n!} \frac{\Gamma(\tilde{\alpha} + n) \Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha}) \Gamma(\tilde{\alpha} + \tilde{\beta} + n)} {}_1F_1(\tilde{\alpha} + n; \tilde{\alpha} + \tilde{\beta} + n; -\tilde{\mu}),
\]
which is a known result [27], where \( n = 0, 1, \cdots \).

6. Conclusion and discussion. From the viewpoint of evolution, any species undergoes a birth-death reaction process. Birth-death reaction modules would be a basis of many important intracellular processes and have important influence on behavior of the entire cellular system. Traditional studies assumed that birth-death processes are Markovian (memoryless), but realistic cases are that reaction events happen often in a non-Markovian (or memorial manner. In this paper, by analyzing a generalized birth-death process where intrinsic waiting times are assumed to follow non-exponential distributions, we have analytically shown that in contrast to Markovianity, non-Markovianity has nontrivial effects beyond our imagination. For example, the non-Markovian effect is equivalent to the introduction of feedback; compared with Markovianity, non-Markovianity can reduce or amplify the noise, depending on the detail of waiting time distributions; and non-Markovianity can
even induce bimodality. These analytical results indicate that the non-Markovian effect is in general nontrivial and cannot be neglected.

The theory established above for birth-death processes can be easily extended to more complex reaction processes on networks. For example, based on the chemical time random walk theory [1], one can easily establish a sgCME for a general reaction process on a network with arbitrary intrinsic waiting-time distributions. For this, a starting point is the following CME expressed in the terms of Laplace transform

\[ s\tilde{P}(n; s) = \tilde{P}(n; 0) + \sum_{i=1}^{L} \tilde{M}_i(s; n - v_i) \tilde{P}(n - v_i; s) - \sum_{i=1}^{L} \tilde{M}_i(s; n) \tilde{P}(n; s), \] (33)

where \( n \) represents the state of the underlying reaction system, and \( M_i(t; n) \) is a memory function for reaction \( i \). Similar to analysis of the above birth-death process, we can prove that \( K_i(n) = \lim_{s \to 0} \tilde{M}_i(s; n) \) always exists and is finite, and can be explicitly expressed by intrinsic waiting-time distributions as shown in this paper. Furthermore, we can analyze the effect of non-Markovianity in terms of probability and statistical index. However, when studying a general process, one should keep caution on the effect of non-Markovianity on the behavior of the entire network since non-Markovianity would induce new dynamics as shown in this paper.

Although having analyzed simple birth-death processes in this paper, we believe that the theory established here and its extension (in particular, the sgCME for a general non-Markovian reaction network) can have broad applications. The power of the sgME can be enhanced by analyzing other examples such as non-Markovian random walks and diffusion on networks [8, 12, 18, 19, 22, 29, 32, 37], and non-Markovian open quantum systems [7]. We expect that our analytical frameworks will be of use for studies of a variety of phenomena in biological and physical sciences, and indeed in other areas where individual-based models with general waiting time distributions and/or delayed interactions are relevant.

Acknowledgments. This work was supported by grants 11931019, 11775314 and 11475273 from Natural Science Foundation of P. R. China.

REFERENCES

[1] T. Aquino and M. Dentz, Chemical continuous time random walks, Phys. Rev. Lett., 119 (2017), 230601.
[2] A. L. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature, 435 (2005), 207–211.
[3] M. Barrio, K. Burrage, A. Leier and T. Tian, Oscillatory regulation of Hes1 discrete stochastic delay modelling and simulation, PLoS Comput. Biol., 2 (2006), e117.
[4] D. Bratsun, D. Volb, L. S. Tsimring and J. Hasty, Delay-induced stochastic oscillations in gene regulation, Proc. Natl. Acad. Sci. U.S.A., 102 (2005), 14593–14598.
[5] T. Brett and T. Galla, Stochastic processes with distributed delays: Chemical Langevin equation and linear noise approximation, Phys. Rev. Lett., 110 (2013), 250601.
[6] A. Corral, Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes, Phys. Rev. Lett., 92 (2004), 108501.
[7] I. De Vega and D. Alonso, Dynamics of non-Markovian open quantum systems, Rev. Modern Phys., 89 (2017), 015001.
[8] J. Delvenne, R. Lambiotte and L. E. C. Rocha, Diffusion on networked systems is a question of time or structure, Nat. Commun., 6 (2015), 1–10.
[9] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 2, John Wiley & Sons, New York, 2008.
[10] C. W. Gardiner, Stochastic Methods: A Handbook for the Natural and Social Sciences, Springer-Verlag, Berlin, 2009.
[11] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comput. Phys.*, **22** (1976), 403–434.

[12] J. P. Gleeson, K. P. O’Sullivan, R. A. Banos and Y. Moreno, Effects of network structure, competition and memory time on social spreading phenomena, *Phys. Rev. X*, **6** (2016), 021019.

[13] T. Guérin, O. Bénichou and R. Voituriez, Non-Markovian polymer reaction kinetics, *Nat. Chem.*, **4** (2012), 568–573.

[14] C. V. Harper, B. Finkenstädt, D. J. Woodcock, S. Friedrichsen, S. Semprini, L. Ashall, et al., Dynamic analysis of stochastic transcription cycles, *PLoS Biol.*, **9** (2011), e1000607.

[15] H. W. Hethcote and P. v. d. Driessche, A SIS epidemic model with variable population size and a delay, *J. Math. Biol.*, **34** (1995), 177–194.

[16] J. C. Jaeger and G. Newstead, *An Introduction to the Laplace Transformation with Engineering Applications*, Methuen & Co., Ltd., London, John Wiley & Sons, Inc., New York, NY, 1949.

[17] T. Jia and R. V. Kulkarni, Intrinsic noise in stochastic models of gene expression with molecular memory, *Phys. Rev. Lett.*, **106** (2011), 058102.

[18] H. H. Jo, J. I. Perotti, K. Kaski and J. Kertész, Analytically solvable model of spreading dynamics with non-Poissonian processes, *Phys. Rev. X*, **4** (2014), 011041.

[19] I. Z. Kiss, G. Röst and Z. Vizi, Generalization of pairwise models to non-Markovian epidemics on networks, *Phys. Rev. Lett.*, **115** (2015), 078701.

[20] N. Kumar, A. Singh and R. V. Kulkarni, Transcriptional bursting in gene expression: Analytical results for general stochastic models, *PLoS Comput. Biol.*, **11** (2015), e1004292.

[21] D. R. Larson, What do expression dynamics tell us about the mechanism of transcription?, *Curr. Opin. Genet. Dev.*, **21** (2011), 591–599.

[22] N. Masuda, M. A. Porter and R. Lambiotte, Random walks and diffusion on networks, *Phys. Rep.*, **716** (2017), 1–58.

[23] A. S. Novozhilov, G. P. Karev and E. V. Koonin, Biological applications of the theory of birth-and-death processes, *Briefings in Bioinformatics*, **7** (2006), 70–85.

[24] E. Pardoux, *Markov Processes and Applications: Algorithms, Networks, Genome and Finance*, Vol. 796, John Wiley & Sons, New York, 2008.

[25] J. Peccoud and B. Ycart, Markovian modeling of gene product synthesis, *Theor. Popul. Biol.*, **48** (1995), 222–234.

[26] J. M. Pedraza and J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression, *Science*, **319** (2008), 339–343.

[27] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, *PLoS Biol.*, **4** (2006), e309.

[28] 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. U.S.A.*, **107** (2010), 22020–22025.

[29] I. Scholtes, N. Wider, R. Pfützner, A. Garas, C. J. Tessone and F. Schweitzer, Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks, *Nat. Commun.*, **5** (2014), 1–9.

[30] A. Schwabe, K. N. Rybakova and F. J. Bruggeman, Transcription stochasticity of complex gene regulation models, *Biophys. J.*, **103** (2012), 1152–1161.

[31] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008), 17256–17261.

[32] M. Starnini, J. P. Gleeson and M. Boguñá, Equivalence between non-Markovian and Markovian dynamics in epidemic spreading processes, *Phys. Rev. Lett.*, **118** (2017), 128301.

[33] P. S. Stumpf, R. C. Smith, M. Lenz, A. Schuppert, F. J. Müller, A. Babtie, T. E. Chan, M. P. Stumpf, C. P. Please, S. D. Howison, F. Arai and B. D. MacArthur, Stem cell differentiation as a non-Markov stochastic process, *Cell Syst.*, **5** (2017), 268–282.

[34] 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** (2011), 472–474.

[35] P. Thomas, N. Popović and R. Grima, Phenotypic switching in gene regulatory networks, *Proc. Natl. Acad. Sci. U.S.A.*, **111** (2014), 6994–6999.

[36] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, 2007.

[37] P. Van Mieghem and R. Van de Bovenkamp, Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks, *Phys. Rev. Lett.*, **110** (2013), 108701.
[38] J. J. Zhang and T. S. Zhou, Promoter-mediated transcriptional dynamics, *Biophys. J.*, **106** (2014), 479–488.

[39] J. J. Zhang and T. S. Zhou, Markovian approaches to modeling intracellular reaction processes with molecular memory, *Proc. Natl. Acad. Sci. U. S. A.*, **116** (2019), 23542–23550.

[40] J. J. Zhang, Q. Nie and T. S. Zhou, A moment-convergence method for stochastic analysis of biochemical reaction networks, *J. Chem. Phys.*, **144** (2016), 194109.

Received October 2019; revised June 2020.

E-mail address: zhangzhq6@mail2.sysu.edu.cn
E-mail address: chenmling@mail2.sysu.edu.cn
E-mail address: zhjiajun@mail.sysu.edu.cn
E-mail address: mcszhtsh@mail.sysu.edu.cn